VOLUME 104 PART B NUMBER 17 SEPTEMBER 1957

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 : INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE · LONDON W.C.2

Price Ten Shillings

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

COUNCIL 1956–1957

President

SIR GORDON RADLEY, K.C.B., C.B.E., Ph.D.(Eng.).

Past-Presidents

SIR JAMES SWINBURNE, Bart., F.R.S.
W. H. ECCLES, D.Sc., F.R.S.
THE RT. HON. THE EARL OF MOUNT EDGCUMBE, T.D.
J. M. DONALDSON, M.C.
PROFESSOR E. W. MARCHANT, D.Sc.
H. T. YOUNG.
SIR GEORGE LEE, O.B.E., M.C.
SIR ARTHUR P. M. FLEMING, C.B.E., D.Eng., LL.D.
J. R. BEARD, C.B.E., M.Sc.
SIR NOEL ASHBRIDGE, B.Sc.(Eng.).
COLONEL SIR A. STANLEY ANGWIN, K.C.M.G., K.B.E., D.S.O., M.C., T.D., D.Sc.(Eng.).
SIR HARRY RAILING, D.Eng.
P. DUNSHEATH, C.B.E., M.A., D.Sc.(Eng.).
SIR VINCENT Z. DE FERRANTI, M.C.
T. G. N. HALDANE, M.A.
PROFESSOR E. B. MOULLIN, M.A., Sc.D.
SIR ARCHIBALD J. GILL, B.Sc.(Eng.).
SIR JOHN HACKING.
COLONEL B. H. LEESON, C.B.E., T.D.
SIR HAROLD BISHOP, C.B.E., B.Sc.(Eng.).
SIR JOSIAH ECCLES, C.B.E., D.Sc.
SIR GEORGE H. NELSON, Bart.

Vice-Presidents

T. E. GOLDUP, C.B.E.
S. E. GOODALL, M.Sc.(Eng.).
WILLIS JACKSON, D.Sc., D.Phil., Dr.Sc.Tech., F.R.S.
SIR HAMISH D. MACLAREN, K.B.E., C.B., D.F.C., LL.D., B.Sc.
G. S. C. LUCAS, O.B.E.

Honorary Treasurer

THE RT. HON. THE VISCOUNT FALMOUTH.

Ordinary Members of Council

PROFESSOR H. E. M. BARLOW, Ph.D., B.Sc.(Eng.).
J. BENNETT.
J. A. BROUGHALL, B.Sc.(Eng.).
C. M. COCK.
A. R. COOPER, M.Eng.
B. DONKIN, B.A.
PROFESSOR J. GREIG, M.Sc., Ph.D.
E. M. HICKIN.
F. J. LANE, O.B.E., M.Sc.
D. MCDONALD, B.Sc.
C. T. MELLING, C.B.E., M.Sc.Tech.
H. H. MULLENS, B.Sc.
A. H. MUMFORD, O.B.E., B.Sc.(Eng.).
W. F. PARKER.
D. P. SAYERS, B.Sc.
G. L. WATES, J.P.
H. WATSON-JONES, M.Eng.
D. B. WELBOURN, M.A.
H. WEST, M.Sc.
J. H. WESTCOTT, B.Sc.(Eng.), Ph.D.
E. L. E. WHEATCROFT, M.A.

Chairmen and Past-Chairmen of Sections

Measurement and Control:
D. TAYLOR, M.Sc., Ph.D.
*W. BAMFORD, B.Sc.

Radio and Telecommunication:
R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).
*H. STANESBY.

Supply:
P. J. RYLE, B.Sc.(Eng.).
*L. DRUCQUER.

Utilization:
H. J. GIBSON, B.Sc.
*D. B. HOGG, M.B.E., T.D.

Chairmen and Past-Chairmen of Local Centres

East Midland Centre:
H. L. HASLEGRAVE, M.A., Ph.D., M.Sc.(Eng.).
*F. R. C. ROBERTS.

North Midland Centre:
W. K. FLEMING.
*F. BARRELL.

North-Western Centre:
T. E. DANIEL, M.Eng.
*G. V. SADLER.

Scottish Centre:
PROFESSOR F. M. BRUCE, M.Sc., Ph.D
*E. WILKINSON, Ph.D., B.Eng.

Mersey and North Wales Centre:
P. D'E. STOWELL, B.Sc.(Eng.).
*PROFESSOR J. M. MEEK, D.Eng.

North-Eastern Centre:
J. CHRISTIE.
*A. H. KENYON.

Northern Ireland Centre:
DOUGLAS S. PARRY.
*MAJOR E. N. CUNLIFFE, B.Sc.Tech.

South Midland Centre
C. J. O. GARRARD, M.Sc.
*H. S. DAVIDSON, T.D.

Southern Centre:
H. ROBSON, B.Sc.
*L. H. FULLER, B.Sc.(Eng.).

Western Centre:
PROFESSOR G. H. RAWCLIFFE, M.A., D.Sc.
*T. G. DASH, J.P.

* Past Chairmen.

MEASUREMENT AND CONTROL SECTION COMMITTEE 1956–1957

Chairman

DENIS TAYLOR, M.Sc., Ph.D.

Vice-Chairmen

H. S. PETCH, B.Sc.(Eng.); J. K. WEBB, M.Sc.(Eng.), B.Sc.Tech.

Past-Chairmen

W. BAMFORD, B.Sc.; M. WHITEHEAD.

Ordinary Members of Committee

J. BELL, M.Sc.
PROFESSOR F. BRAILSFORD, Ph.D., B.Sc.(Eng.).
D. EDMUNDSON, B.Sc.(Eng.).
W. S. ELLIOTT, M.A.
PROFESSOR K. A. HAYES, B.Sc.(Eng.).
M. KAUFMANN.
W. C. LISTER, B.Sc.
R. S. MEDLOCK, B.Sc.
C. RYDER.
R. H. TIZARD, B.A.
PROFESSOR A. TUSTIN, M.Sc.
M. V. WILKES, M.A., Ph.D.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
PROFESSOR J. GREIG, M.Sc., Ph.D. (representing the Council).
C. H. W. LACKEY, B.Sc. (representing the North-Eastern Radio and Measurements Group.
F. BEECH (representing the North-Western Measurement and Control Group).
D. O. CLAYDEN (nominated by the National Physical Laboratory).

RADIO AND TELECOMMUNICATION SECTION COMMITTEE 1956–57

Chairman

R. C. G. WILLIAMS, Ph.D., B.Sc.(Eng.).

Vice-Chairmen

J. S. MCPETRIE, Ph.D., D.Sc.
G. MILLINGTON, M.A., B.Sc.

Past-Chairmen

H. STANESBY.
C. W. OATLEY, O.B.E., M.A., M.Sc.

Ordinary Members of Committee

A. J. BIGGS, Ph.D., B.Sc.
W. J. BRAY, M.Sc.(Eng.).
H. A. M. CLARK, B.Sc.(Eng.).
C. W. EARP, B.A.
V. J. FRANCIS, B.Sc.
E. V. D. GLAZIER, Ph.D.(Eng.), B.Sc.
W. ROSS, M.A.
L. RUSHWORTH, M.B.E., B.Sc.
T. B. D. TERRONI, B.Sc.
A. M. THORNTON, B.Sc.
D. R. TURNER, M.Eng.
F. WILLIAMS, B.Sc.

And

The President (*ex officio*).
The Chairman of the Papers Committee.
PROF. H. E. M. BARLOW, Ph.D., B.Sc.(Eng.) (representing the Council).
E. H. COOKE-YARBOROUGH (Co-opted Member).
J. G. YATES, M.A. (representing the Cambridge Radio and Telecommunication Group).
J. MOIR (representing the South Midland Radio and Telecommunication Group).
A. E. TWYCROSS (representing the North-Eastern Radio and Measurements Group).
A. C. NORMINGTON, B.Sc.(Eng.) (representing the North-Western Radio and Telecommunication Group).

The following nominees of Government Departments:
Admiralty: CAPTAIN A. J. B. NAISH, M.A., R.N.
Air Ministry:
Department of Scientific and Industrial Research: B. G. PRESSEY, M.Sc.(Eng.), Ph.D.
Ministry of Supply: BRIG. J. D. HAIGH, O.B.E., M.A.
Post Office: CAPTAIN C. F. BOOTH, C.B.E.
War Office: COL. E. I. E. MOZLEY, M.A.

Secretary

W. K. BRASHER, C.B.E., M.A., M.I.E.E.

Deputy Secretary

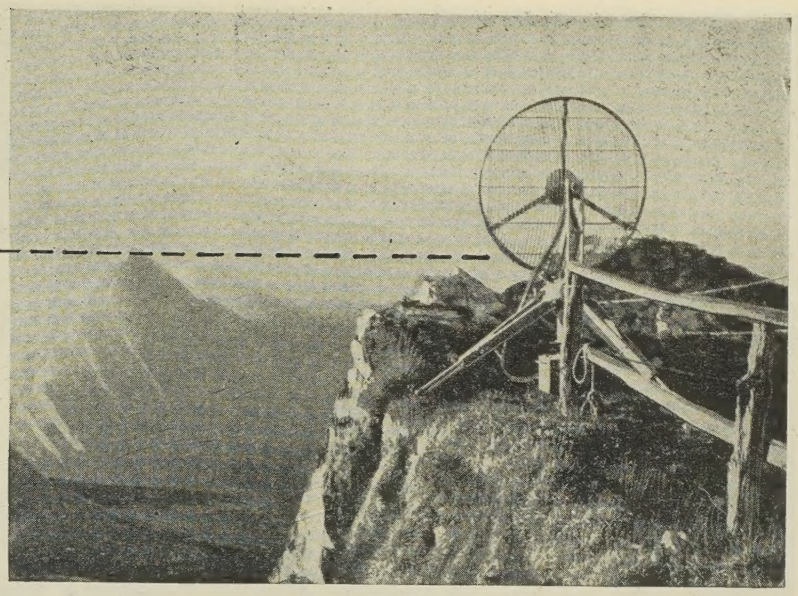
F. JERVIS SMITH, M.I.E.E.

Assistant Secretary

F. C. HARRIS.

Editor-in-Chief

G. E. WILLIAMS, B.Sc.(Eng.), M.I.E.E.

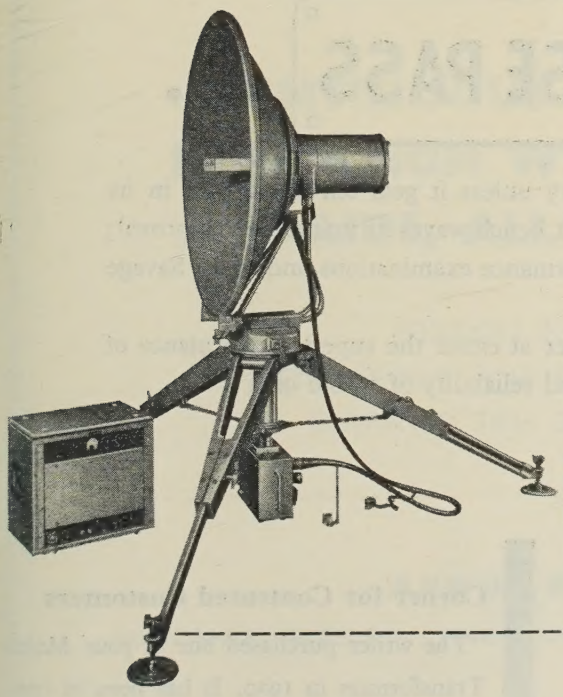


One of the E.M.I. Television Microwave Relay Links in operation in the Swiss Alps forms part of the Eurovision Network.

E·M·I

TELEVISION Relay Equipment

E.M.I. High Power Television Microwave Links are designed to provide a high quality video link between an outside broadcast unit and the main T.V. station or for projects such as the Eurovision Network. They are suitable for 405, 525 and 635 line systems, are portable and of robust construction. Range over unobstructed path at least 40 miles.



TYPE ML6B (3 Watt Klystron)

- Operating Frequency 4,400-4,800 Mc/s per sec.
- Available with both sound and vision channels.

TYPE ML4A (2 Watt Klystron)

- Operating Frequency 6,875-7,300 Mc/s per sec.
- Suitable for colour as well as monochrome transmissions.

E.M.I... PIONEERS OF THE WORLD'S FIRST PUBLIC HIGH DEFINITION TELEVISION SERVICE IN 1936 . . . E.M.I. PATENTED TELEVISION INVENTIONS ARE EXTENSIVELY USED UNDER LICENCE.

OTHER E.M.I. DEVELOPMENTS FOR TELEVISION TRANSMISSION

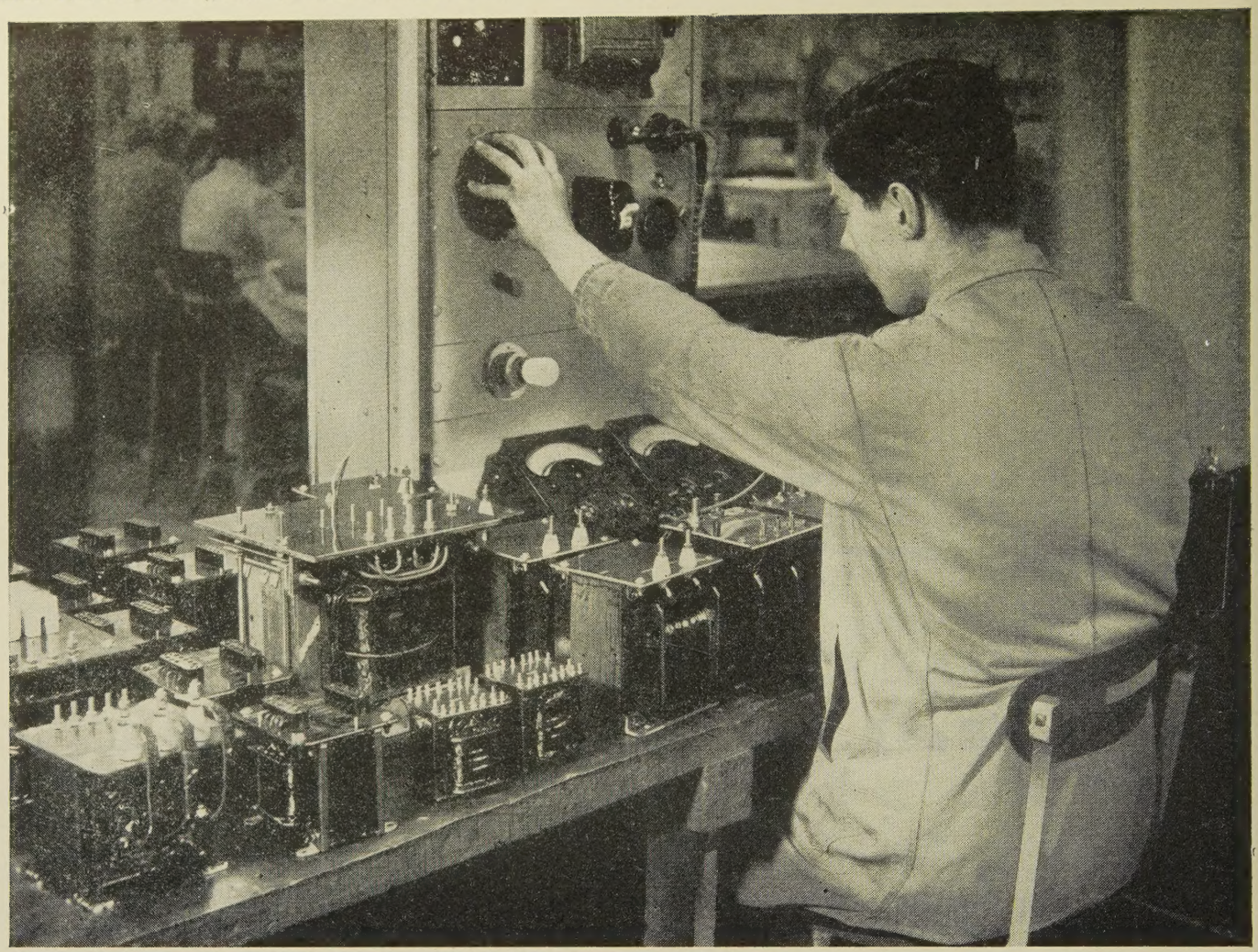
New H. F. Transmitters, Camera Channels and Associated Equipments. Monochrome, Photo-Conductive and Flying Spot Film Channels and Colour Flying Spot Film Channels. Also Industrial Television.

For particulars of the above-mentioned equipments write to:—

E.M.I. ELECTRONICS LTD., VALVE DIVISION, HAYES, MIDDLESEX, ENGLAND

Tel: SOUthall 2468 Ext. 316

EE93



ON TEST-PLEASE PASS . . .

. . . but no Massicore Transformer is ever passed for duty unless it gets ten out of ten in its performance examination and until every needle on the test bench waves its unqualified approval; and Massicore Transformers must pass much higher performance examinations and more Savage tests than they are ever likely to meet in service.

Many customers of long-standing still express their wonder at either the superb performance of a new Massicore Instrument or the stalwart consistency and reliability of an old one.

May we amaze *you*, new customer?

MASSICORE SAVAGE

SAVAGE TRANSFORMERS LTD

NURSTEED ROAD · DEVIZES · WILTSHIRE · Tel: Devizes 932

Corner for Contented Customers

"The writer purchased one of your Mains Transformers in 1930. It has been in continuous service since that date and when last seen by me was still functioning perfectly . . ."

C. E. Ltd. Southampton

TP/55

MARCONI

MILITARY AND CIVIL

airborne

DOPPLER NAVIGATORS

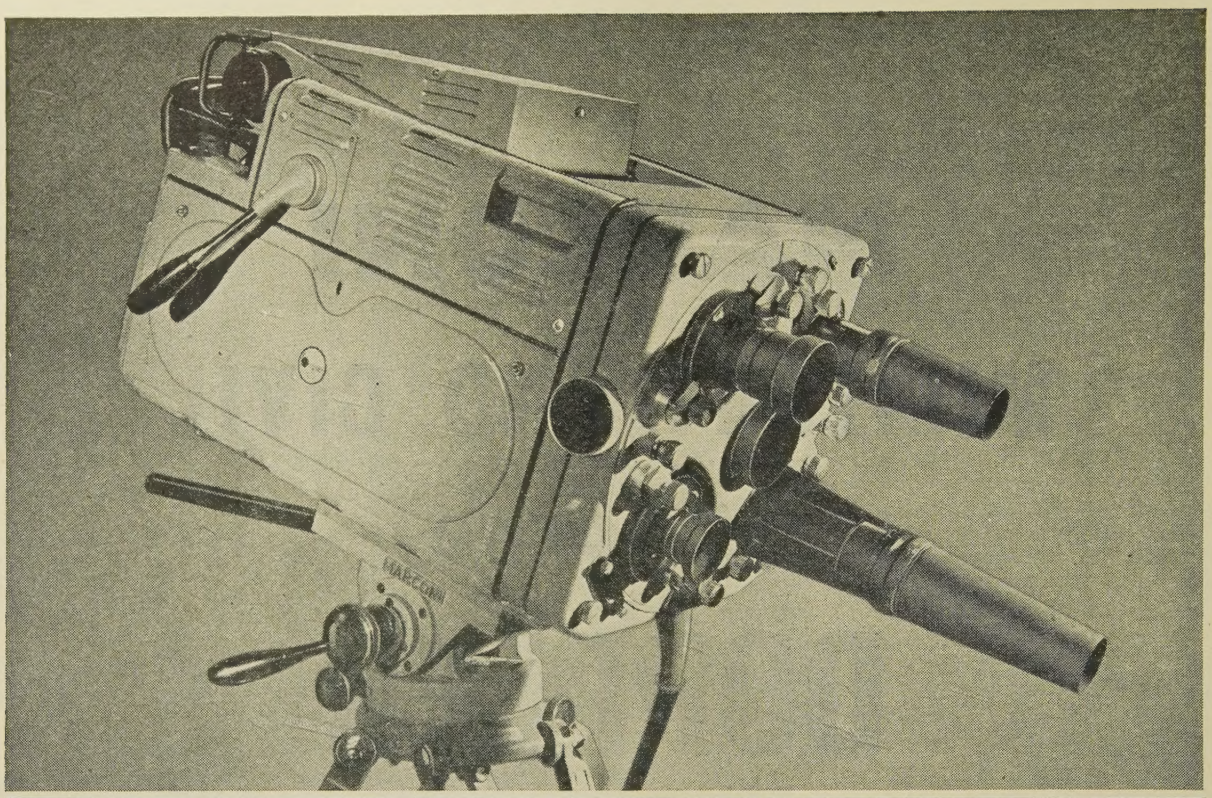
CONTINUOUS AUTOMATIC POSITION INDICATION WITHOUT GROUND-BASED AIDS, ALL OVER THE WORLD

POSITION IN LATITUDE AND LONGITUDE
DISTANCE RUN AND DISTANCE TO GO • WIND VELOCITY
ESTIMATED TIME OF ARRIVAL • TRACK GUIDANCE

Write for details to

AERONAUTICAL DIVISION

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ENGLAND



Marconi Camera Channels

IMAGE ORTHICON CAMERA Type BD808 (illustrated)

Features

- Uses either 3″ or 4½″ Image Orthicons.
- Designed for ease of servicing, excellent accessibility and plug-in sub-units.
- Four position turret will carry any combination from 2-inch to 40-inch lenses. 80-inch and zoom lenses may also be used.
- Viewfinder can be tilted up or down to give the most comfortable viewing position.
- Camera Control Unit may be used with 10″ picture tube and 3″ waveform tube, or with 14″ picture tube and 5″ waveform tube.
- Remote control of light intensity by variable graded filter.
- Optional remote control of focus and turret. Optional semi-automatic alignment circuit.
- Built-in turret for neutral density and colour filters.
- Full range of accessories available for both studio and outside broadcast roles.

BROADCAST VIDICON CAMERA Type BD864

The most recent addition to the Marconi range of Television Equipment.

Features

- Compact, easily operated by one man. The camera has integral viewfinder with 7″ tube and 2¾″ waveform monitor and includes all operational controls.
- Channel consists of Camera and Power Supply only—but optional Remote C.C.U. and Monitor position available.
- Use of close-tolerance double-triodes in all valve circuits except one and printed wiring assemblies ensures great reliability.
- Rapid semi-automatic beam alignment, built-in aperture correction and gamma correction circuits are provided. Designed to make the best use of any of the present Vidicon tubes and with ample flexibility to deal with foreseeable developments.
- 4-position turret with positive location takes wide range of fixed and zoom lenses.

MARCONI

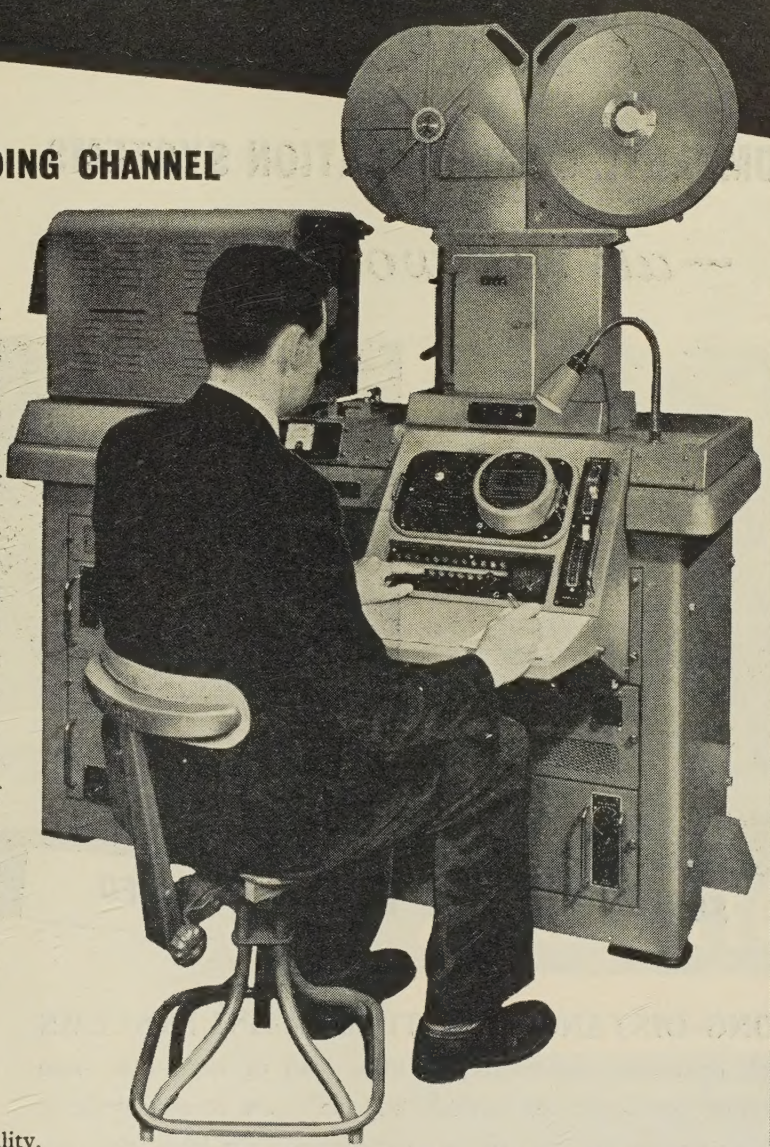
COMPLETE SOUND BROADCASTING AND TELEVISION SYSTEMS

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX

LB9

MARCONI TELEVISION RECORDING EQUIPMENT

TYPE BD679 RECORDING CHANNEL



16mm Film presents a most economical and flexible means of recording Television pictures since film costs are low, developing and printing techniques are advanced and a wide range of fine grain film stocks are readily available. Projection, editing, viewing, dubbing and handling are all easily carried out with standard equipment. The Fast Pull-down technique presents great advantages over other systems of recording on film but the mechanical difficulties of moving the film in the short period of frame blanking have so far prevented the employment of the technique. These problems have now been successfully overcome by Marconi's.

Features

Exceptionally high picture quality.

Specially developed gearbox enables Fast Pull-down technique to be employed. F.P.D. Mechanism has given over 3,000 hours trouble-free operation.

Pull-down time adjustable, normally set at 2 milliseconds permitting recording of fully interlaced picture. Simple single-lens optical system avoids loss of contrast.

Sound can be recorded on optical track, magnetic stripe or separate synchronous magnetic track.

Conveniently placed input selector switches, monitor and level controls. Sound/Vision cueing device incorporated.

Recording can be made on positive or negative stock of a wide variety, either direct positive, direct negative or reversal. Magazines hold 2,400 ft. (2,000 ft. magnetic stripe) or film may be fed directly into a rapid processor.

MARCONI

COMPLETE SOUND AND TELEVISION BROADCASTING SYSTEMS

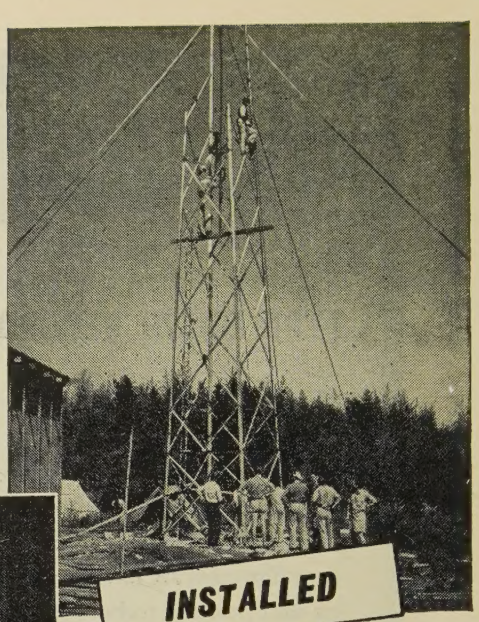
MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX

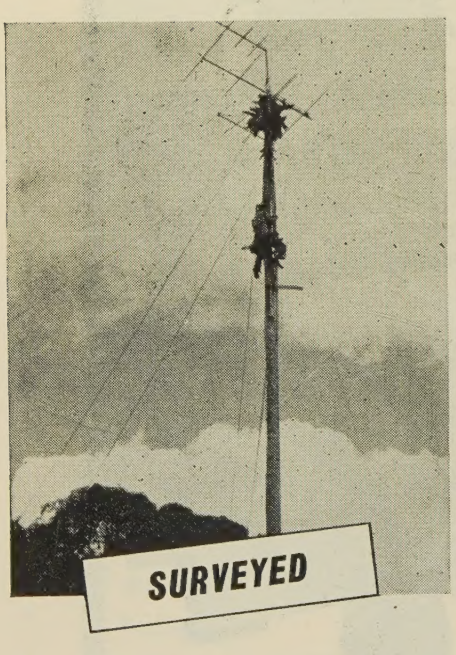
LB10

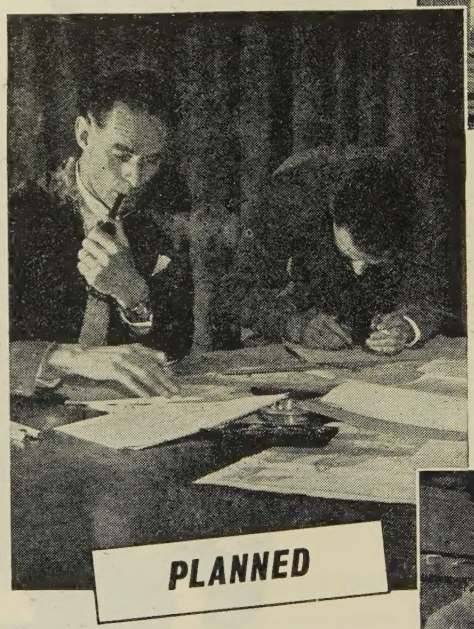
MARCONI

COMPLETE COMMUNICATION SYSTEMS

—*all the world over*

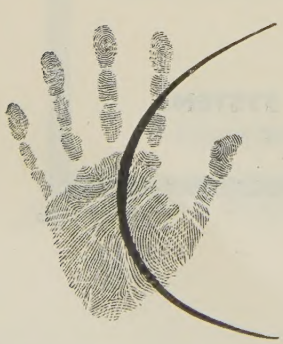






LONG-DISTANCE H.F. TELEGRAPH SYSTEMS

High Frequency systems form a major part of world-wide radio telegraph communication services. Marconi's have recently designed new equipment for such systems incorporating the latest electronic developments to save time and labour, reduce operating costs and eliminate faults. The company is unique in the resourcefulness and skill it can bring to the complete engineering of a system from the surveying stage onwards to the maintenance after it has been installed, and the training of the staff to operate it at maximum efficiency.

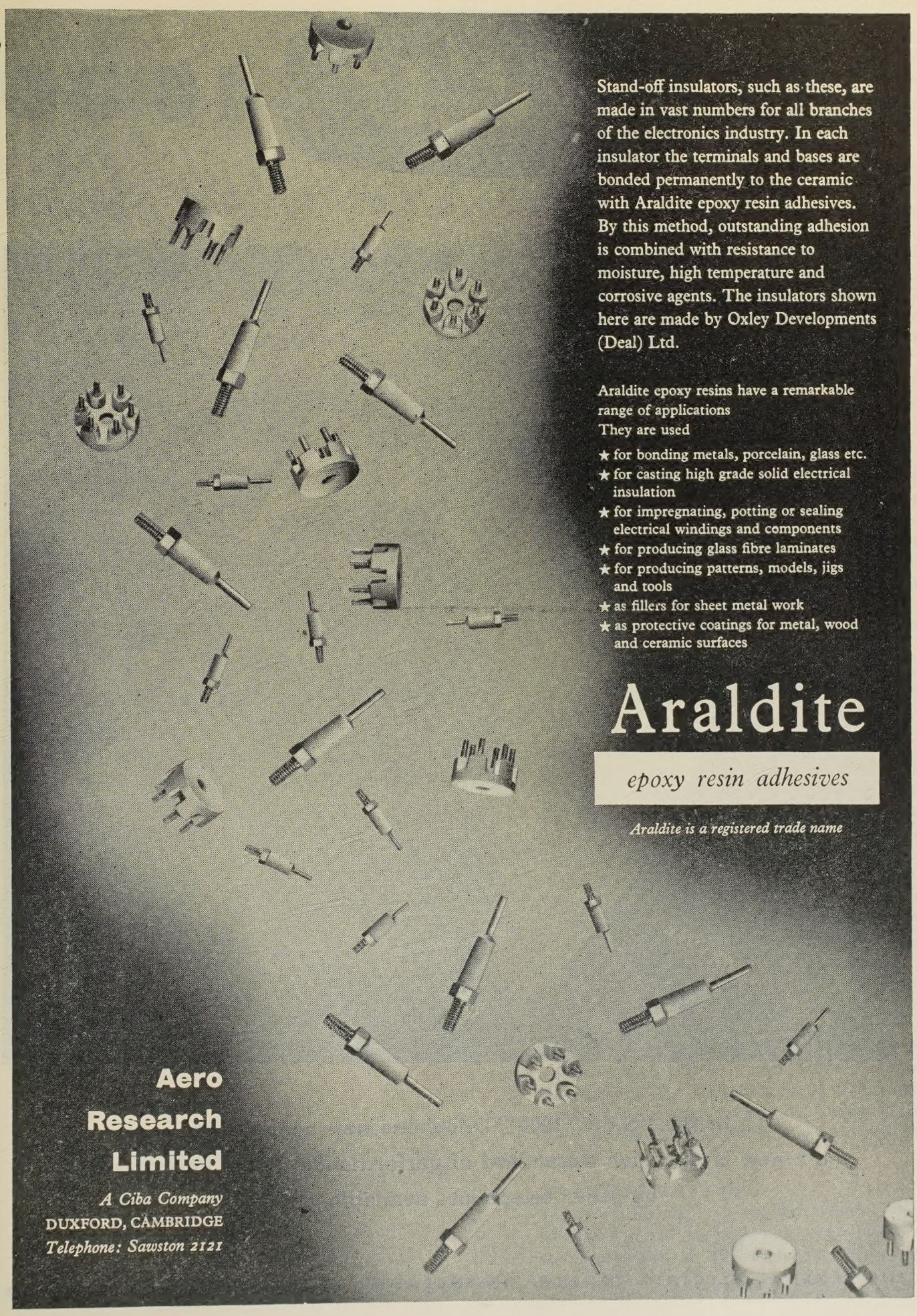


The Lifeline of Communication is in experienced hands

MARCONI

Complete Communication Systems

MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX


Stand-off insulators, such as these, are made in vast numbers for all branches of the electronics industry. In each insulator the terminals and bases are bonded permanently to the ceramic with Araldite epoxy resin adhesives. By this method, outstanding adhesion is combined with resistance to moisture, high temperature and corrosive agents. The insulators shown here are made by Oxley Developments (Deal) Ltd.
Araldite epoxy resins have a remarkable range of applications
They are used
★ for bonding metals, porcelain, glass etc.
★ for casting high grade solid electrical insulation
★ for impregnating, potting or sealing electrical windings and components
★ for producing glass fibre laminates
★ for producing patterns, models, jigs and tools
★ as fillers for sheet metal work
★ as protective coatings for metal, wood and ceramic surfaces
Araldite
epoxy resin adhesives
Araldite is a registered trade name
Aero Research Limited
A Ciba Company
DUXFORD, CAMBRIDGE
Telephone: Sawston 2121

The G.E.C. 1000

In the "G.E.C. 1000" Telephone new components in a new circuit give a standard of performance that is unsurpassed by other telephones available today

THE GENERAL ELECTRIC CO. LTD. OF ENGLAND · TELEPHONE RADIO AND

Telephone

The appearance has been redesigned to blend with modern styles of decoration, whilst retaining the dignity essential to office furnishings.

PERFORMANCE

The rocking-armature receiver has an excellent frequency response and an output 10dB higher than its predecessors. Part of this improvement is transferred by the induction coil to the sending side, giving a resultant improvement in performance of 6dB on sending and 4dB on receiving.

This improved performance means that the new G.E.C. telephones will operate over a local-line loop of 1,120 ohms of 6½-lb cable (0·5mm conductor) with a performance equal to that of previous telephones operating over a local-line loop of 660 ohms, i.e. a local line may be extended by 70%. Alternatively, the new telephones will operate over a length of 4-lb. cable (0·4 mm) with a performance equal to that of a previous telephone operating over the same length of 10-lb. cable.

Despite the increased level of performance, the maximum amount of sidetone suppression has been retained.

COLOUR

In addition to the normal black instrument, a range of two-tone telephones can be supplied in which the case is coloured and all other parts, including the dial, are black. The range of colours is red, green and ivory. The two-tone telephones have the following advantages over the all-colour telephones.

More delicate shades can be used since shade matching is eliminated.

The number of spare parts required to be held by an Administration using more than one colour of instrument is greatly reduced.

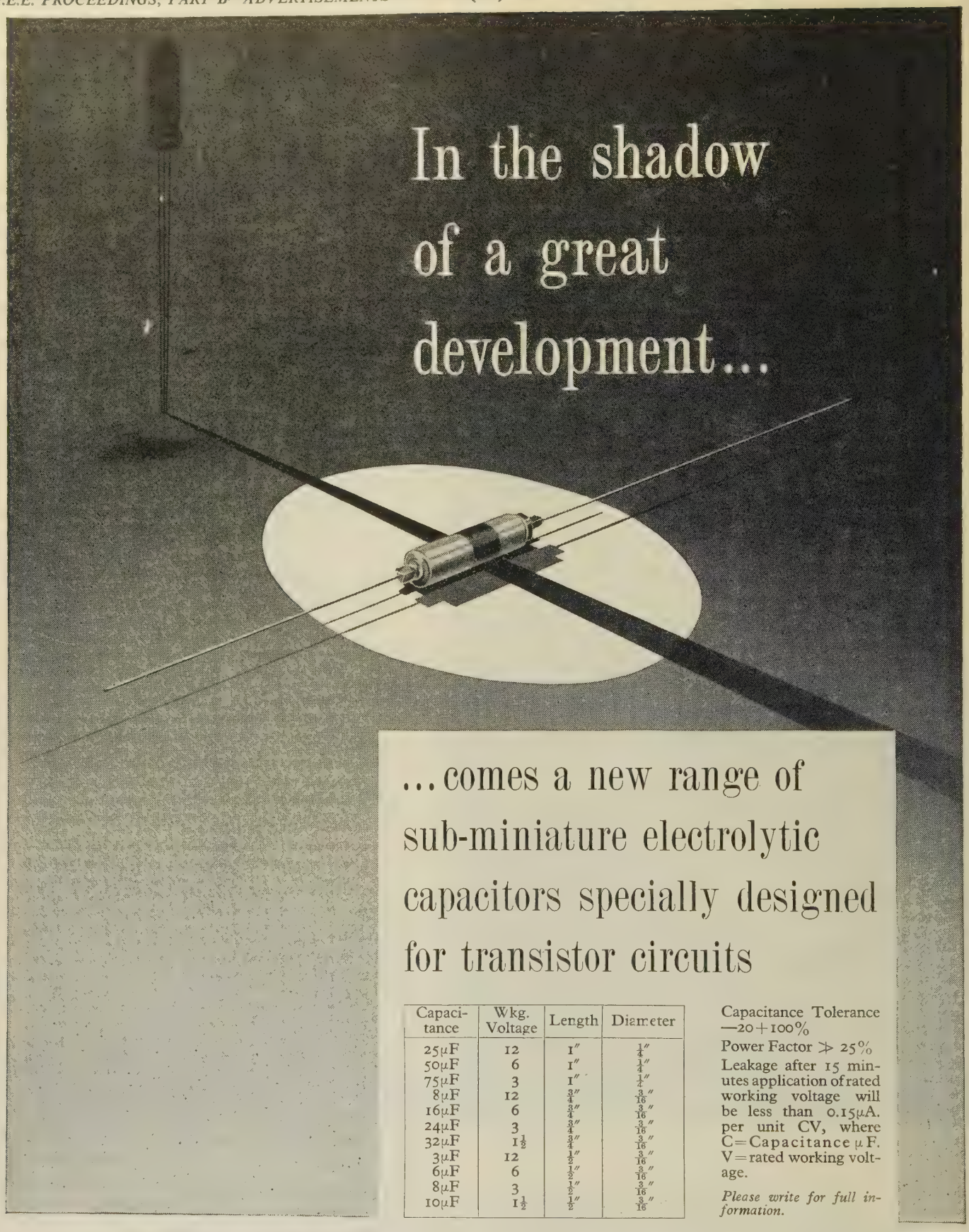
APPEARANCE

The pleasing appearance is achieved by blending a number of curved lines and surfaces to form outlines for the case and handset that are in complete harmony with one another. The camber of the sloping front houses the dial in the automatic telephone and the dial dummy in the C.B. set. The incline of the telephone front is such that the telephone is easy to use, and the dial numbers easy to see, whether the user is in a standing or sitting position.

A curved insert is fitted in each side of the case in front of the cradle to provide finger-tip grips for lifting and carrying the instrument. When resting in the cradle, the curved handset gives the telephone a domed silhouette in accord with the remainder of the instrument. The increased curvature of the handset over previous types gives greater comfort to the user, and tilts the transmitter to a more sensitive position.

TROPICALISATION

The three G.E.C. features—special insulation, ventilation, and protection against moisture and insects—are incorporated in all telephones supplied to tropical areas.



In the shadow of a great development...

...comes a new range of sub-miniature electrolytic capacitors specially designed for transistor circuits

Capacitance	Wkg. Voltage	Length	Diameter
25μF	12	1″	$\frac{1}{4}$″
50μF	6	1″	$\frac{1}{4}$″
75μF	3	1″	$\frac{1}{4}$″
8μF	12	$\frac{3}{4}$″	$\frac{3}{16}$″
16μF	6	$\frac{3}{4}$″	$\frac{3}{16}$″
24μF	3	$\frac{3}{4}$″	$\frac{3}{16}$″
32μF	$1\frac{1}{2}$	$\frac{3}{4}$″	$\frac{3}{16}$″
3μF	12	$\frac{1}{2}$″	$\frac{3}{16}$″
6μF	6	$\frac{1}{2}$″	$\frac{3}{16}$″
8μF	3	$\frac{1}{2}$″	$\frac{3}{16}$″
10μF	$1\frac{1}{2}$	$\frac{1}{2}$″	$\frac{3}{16}$″

Capacitance Tolerance —20+100%

Power Factor ≯ 25%

Leakage after 15 minutes application of rated working voltage will be less than 0.15μA. per unit CV, where C=Capacitance μF. V=rated working voltage.

Please write for full information.

DUBILIER

DUBILIER CONDENSER CO. (1925) LTD., DUCON WORKS, VICTORIA ROAD, NORTH ACTON, LONDON W.3

Telephone: ACOrn 2241

Telegrams: Hivoltcon Wesphone London.

DN 187.

S. HERBERT

SALMON receive communication of approaching objects through the sensory organs situated just below the skin immediately under the line of different hue along their sides. These sensory nerves, at the forward end of the fish, spread out to form the organ which serves the fish as an ear.

SALMONIDEA SALAR

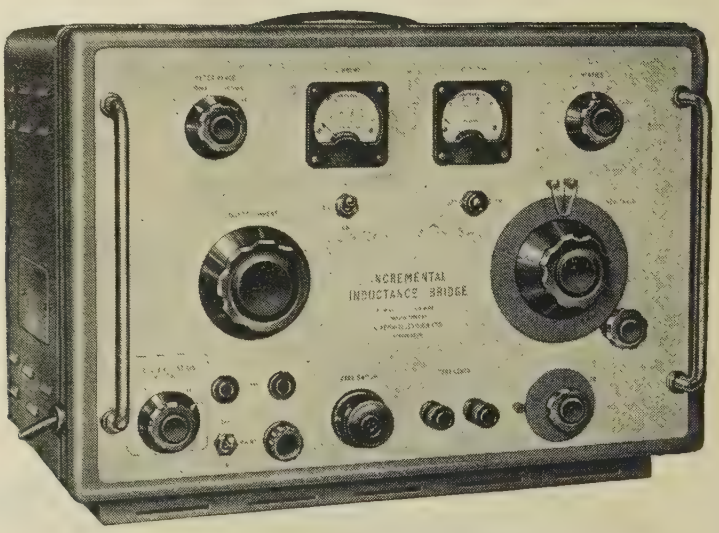
...but when it's telecommunications it's

Automatic Telephone & Electric Co. Ltd

LONDON AND LIVERPOOL

AT1841

INCREMENTAL INDUCTANCE BRIDGE



Designed to measure the value of iron cored chokes and similar inductors in the range 0·01H to 1000H of Q value not less than 2.

Provision is made for passing any current up to 1 Amp d.c. through the winding and selectable a.c. excitation voltages of 1, 2, 5, 10 and 20V r.m.s. are provided.

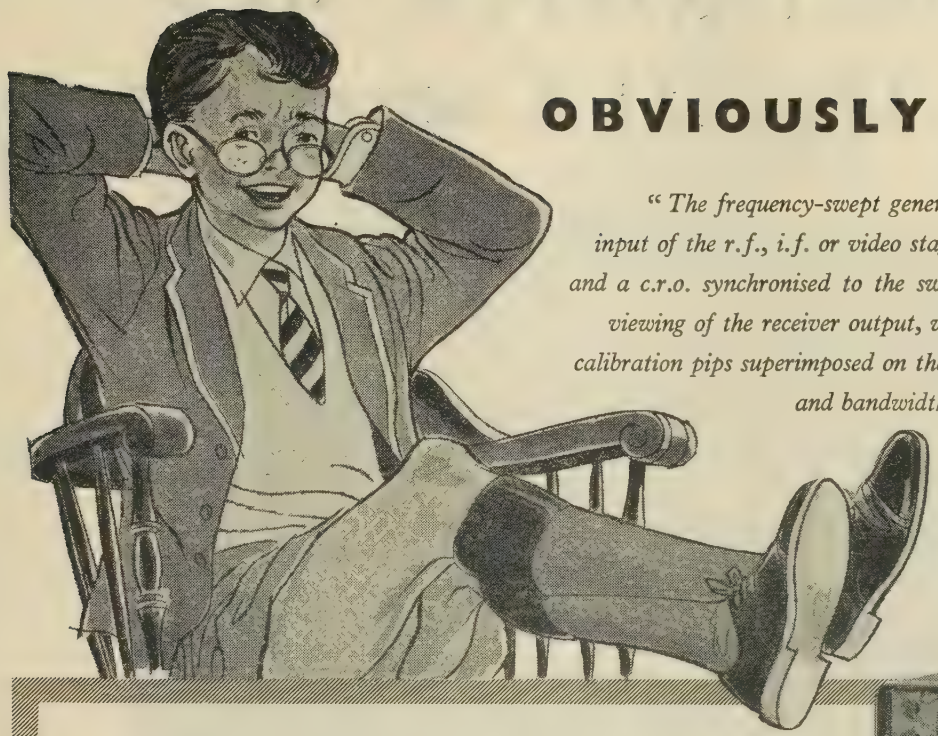
Full technical information is available on request.

CINEMA TELEVISION LTD

A COMPANY WITHIN THE RANK ORGANISATION LIMITED

WORSLEY BRIDGE ROAD · LONDON · S.E.26
HITHER GREEN 4600

SALES AND SERVICING AGENTS:
Hawnt & Co. Ltd., 59 Moor St. Birmingham, 4
Atkins, Robertson & Whiteford Ltd., Industrial Estate, Thornliebank, Glasgow
F. C. Robinson & Partners Ltd., 122 Seymore Grove, Old Trafford, Manchester, 16



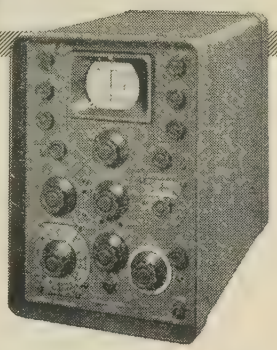
OBVIOUSLY . . .

" The frequency-swept generator connected to the input of the r.f., i.f. or video stages of the receiver, and a c.r.o. synchronised to the sweep frequency, permits viewing of the receiver output, while internally generated calibration pips superimposed on the display enable tuning and bandwidth adjustments to be made with precision."

THAT'S the Marconi V.H.F. Alignment Oscilloscope in a nutshell. If he'd had any breath left our young student might have added that other important applications include the adjustment of discriminators in f.m. receivers and the matching of aerials to transmission lines.

However, the lad's penetrating observation will have revealed to those who are no less observant that here is another important Marconi instrument.

If you don't know as much about it as he does it's time you did. Write and ask for the full facts and we'll send you an informative leaflet post-haste.



MARCONI V.H.F. ALIGNMENT OSCILLOSCOPE
Type TF 1104

For use with television and f.m. receivers. Frequency-swept output and visual display on built-in c.r.t. facilitates rapid evaluation and alignment without ancillary equipment. R.F. ranges: v.h.f. bands I. II and III. I.F. range: 10 to 40 Mc/s. V.F. range: 5 kc/s to 10 Mc/s. Frequency sweep: 10 Mc/s max., with marker pulses at 0·5, 1·0, or 5 Mc/s intervals.

MARCONI INSTRUMENTS

AM & FM SIGNAL GENERATORS · AUDIO & VIDEO OSCILLATORS · FREQUENCY METERS · VOLTMETERS POWER METERS · DISTORTION METERS · FIELD STRENGTH METERS · TRANSMISSION MONITORS DEVIATION METERS · OSCILLOSCOPES, SPECTRUM & RESPONSE ANALYSERS · Q METERS & BRIDGES

MARCONI INSTRUMENTS LTD · ST. ALBANS · HERTFORDSHIRE · TELEPHONE: ST. ALBANS 56161

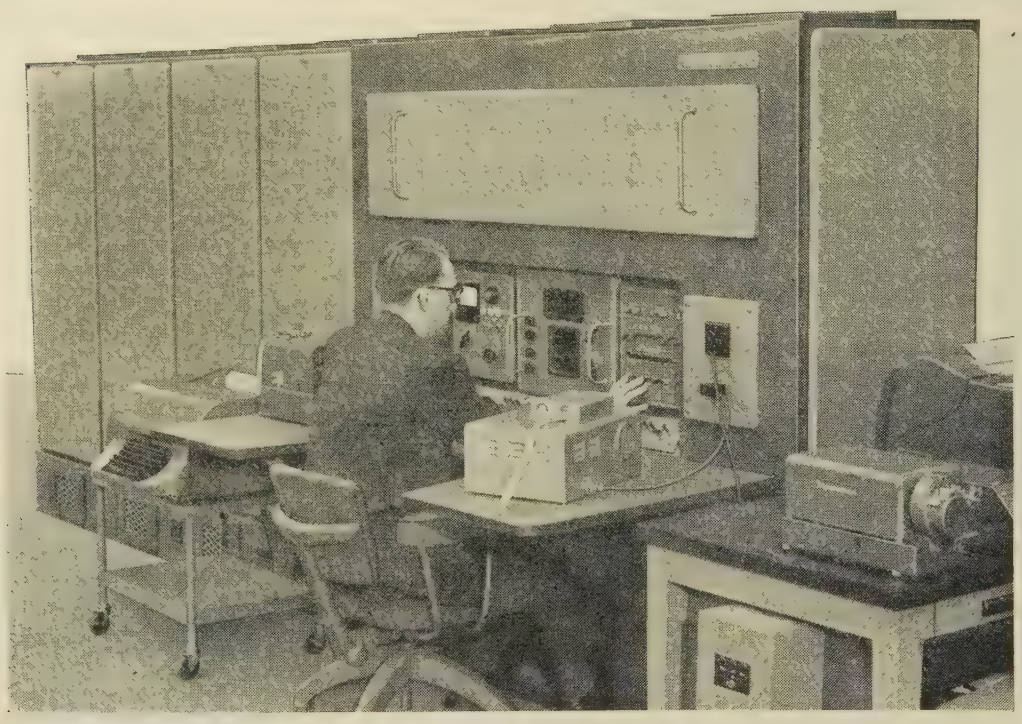
London and the South: Marconi House, Strand, London, W.C.2. Tel: COVent Garden 1234

Midlands: Marconi House, 24 The Parade, Leamington Spa. Tel: 1408 *North:* 30 Albion Street, Kingston-upon-Hull. Tel: Hull Central 16347

WORLD-WIDE REPRESENTATION

TC 104

The METROVICK 950 Digital Computer



FOR MORE RELIABILITY IN COMPUTING

This new computer incorporates :—

FULLY TRANSISTORISED LOGICAL CIRCUITS
PLUG-IN PRINTED CIRCUIT BOARDS

These features in the Metrovick 950 make for—

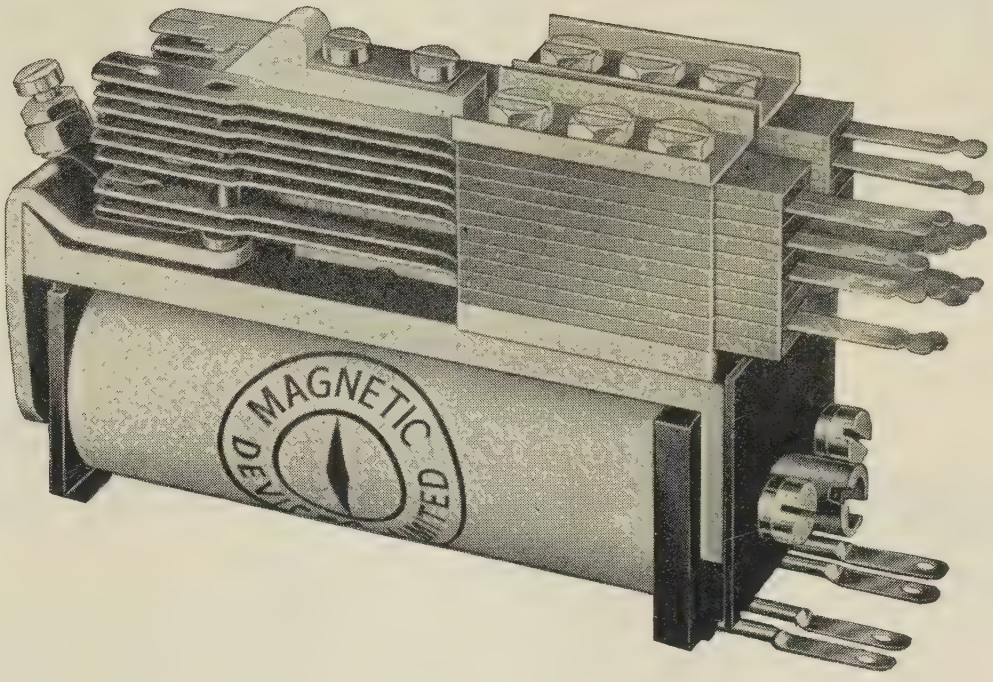
- RELIABILITY
- EASY SERVICING
- COMPACTNESS
- LOW POWER CONSUMPTION
- LOW INITIAL COST

Write for booklet No. SP 7655/1

METROPOLITAN-VICKERS
ELECTRICAL CO LTD · TRAFFORD PARK · MANCHESTER 17

An A.E.I. Company

series 305 relay



P.O. 3000 RELAY

This Relay is the well known P.O. 3000 Relay and can be supplied with coils wound for standard voltages up to 250 volts D.C.

Contact assemblies are available up to six pole changeover and alternative rivets can be supplied to suit varying duties. The Series 305 Relay can be slugged for make or break action and coils can be vacuum impregnated for tropical and humid conditions.

Magnetic Devices LTD

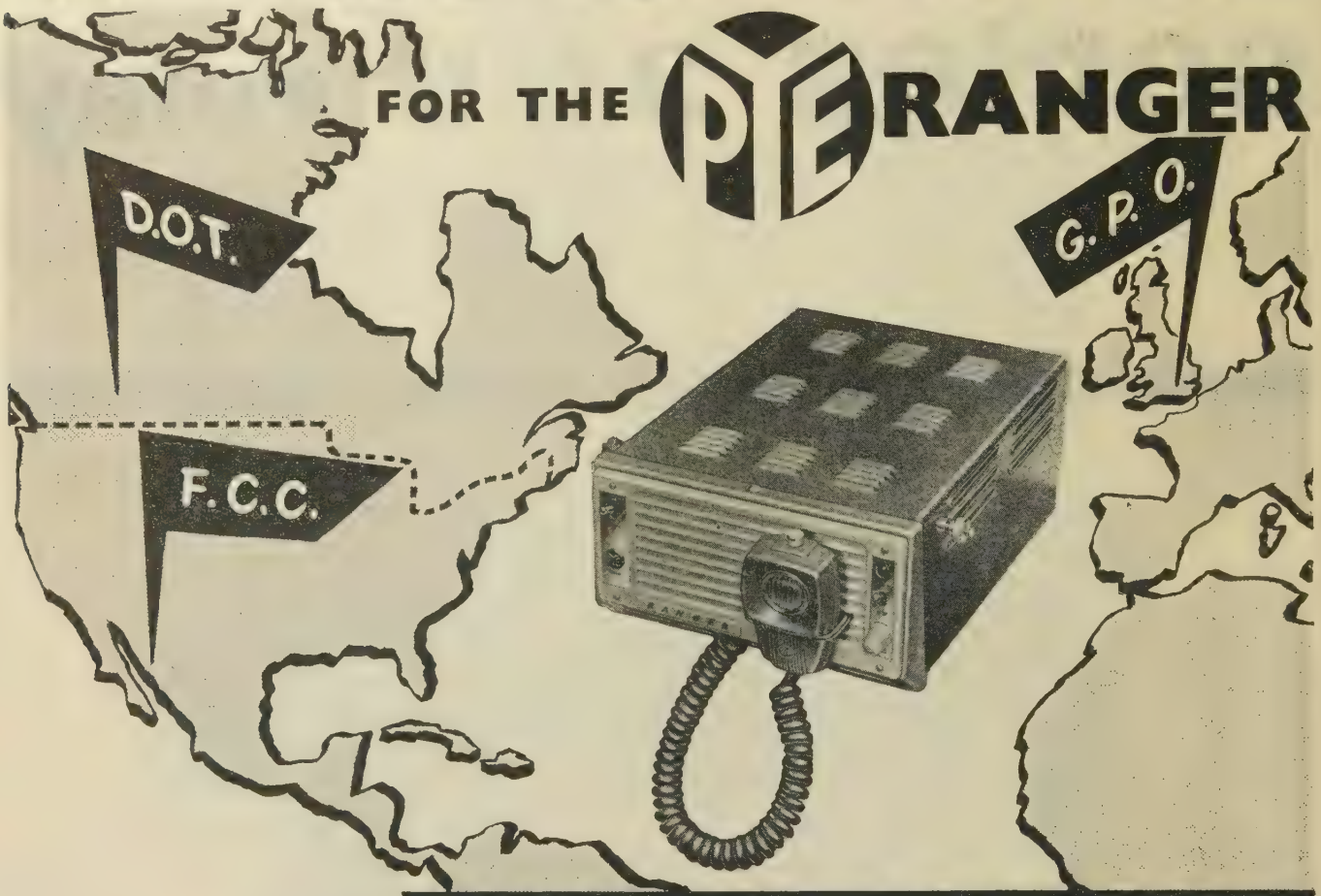
A.I.D. & A.R.B. approved

EXNING ROAD · NEWMARKET · SUFFOLK

Telephone: Newmarket 3181/2/3 Telegrams: MAGNETIC Newmarket.

WORLD APPROVAL

FOR THE PYE RANGER



and 25 Kc/s Channel Spacing

Pye Ranger V.H.F. equipment has received approval from the British G.P.O. for land, marine and international marine applications employing A.M. or F.M. systems, type approval from the Canadian D.O.T., and type acceptance by the F.C.C. of the United States of America.

No other company holds so many approvals for this type of equipment. Ranger equipment now covers every conceivable requirement, and will continue to do so for many years to come.

Ranger equipment now in full production is designed for all channel spacings including 20 and 25 kc/s. for frequency ranges from 25 to 174 Mc/s., for power ranges up to 1 kilowatt and for A.M. or F.M. modulation. This range has been designed to expand the application of Pye Radio-Telephones already in constant use all over the world.

No matter what your V.H.F. requirements, Pye Telecommunications Ltd., can fulfil them. Your inquiries are invited.

Telecommunications

CAMBRIDGE ENGLAND

Pye Telecommunications distributors in 91 countries ensure trouble-free service

PYE TELECOMMUNICATIONS LIMITED · CAMBRIDGE · ENGLAND

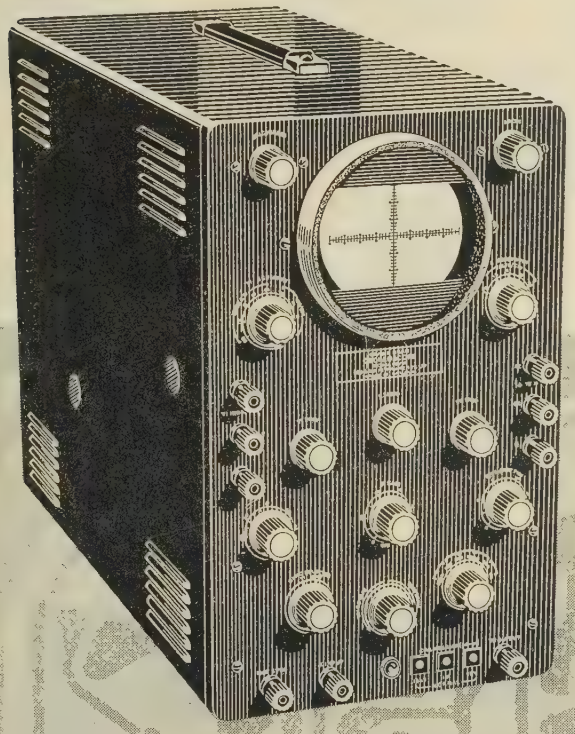
Telephone: Teversham 3131

Cables: Pyetelecom, Cambridge

PERFORMANCE ASSURANCE WITH

COSSOR

PRINTED CIRCUITS



Model 1071K ***Double Beam*** *Kit Oscilloscope*
List Price £57.10.0

AN INSTRUMENT RANGE IN KIT FORM

Q. *Why has Cossor Instruments decided upon this innovation?*

A. To make available a range of first-class measuring instruments **at a considerable saving in cost to the Buyer.**

Q. *Are Kit instruments inferior in performance to their Factory-built equivalents?*

A. Certainly not. If assembled and wired exactly in accordance with the Manual of Instructions.

Q. *A certain skill must, surely, be required to build these instruments?*

A. None beyond the ability to use a small soldering iron.

Q. *How can a performance specification be maintained without setting up with test equipment?*

A. Largely by the use of PRINTED CIRCUITS which allow no interference with the layout of critical parts of the circuit.

Q. *How many Kit instruments are at present available?*

A. Three. Two Oscilloscopes, a Single-Beam and a Double-Beam, and a Valve Voltmeter. Others will follow shortly.

Q. *Could I have more information on these interesting instruments?*

A. With the greatest of pleasure. Just write to:

COSSOR INSTRUMENTS LIMITED

The Instrument Company of the Cossor Group

COSSOR HOUSE · HIGHBURY GROVE · LONDON, N.5

Telephone: CANonbury 1234 (33 lines) Telegrams: Cossor, Norphone, London Cables: Cossor, London

VALVES *for* INDUSTRY

Power rectifiers
High vacuum rectifiers
Transmitting valves
Valves for R.F. heating
Thyratrons
Voltage Stabilisers

The new brochure published by English Electric Valve Co. Ltd., gives data of the widest range of valves for industry, communications, radar and broadcasting made in Great Britain. A copy of this publication will be sent on request.

Magnetrons
Klystrons
Travelling wave tubes
Television camera tubes
Cathode ray tubes
Transistors

'ENGLISH ELECTRIC'

ENGLISH ELECTRIC VALVE CO. LTD. E *Chelmsford, England*
Telephone: Chelmsford 3491

AP/73

First in the field

GOLTOP

POWER TRANSISTORS

available NOW in commercial quantities

These have been in regular quantity production for the past two years, and have proved themselves reliable and stable in a *variety* of applications. They are admirably suitable for all forms of DC to DC or DC to AC Converters, High Power portable Amplifiers and Public Address Equipment. "GOLTOP" Power Transistors are the first to be offered for immediate delivery in quantity. Representing the latest developments in semi-conductor technique for power applications, these entirely British-made p-n-p Germanium Junction Transistors will open up entirely new fields to designers of industrial, commercial and military equipment.

Available in 6 TYPES, all for 10-watts power dissipation:
V15/10P. V15/20P. V15/30P. for 15 volts max.
V30/10P. V30/20P. V30/30P. for 30 volts max.

Maximum Collector Power Dissipation (DC or Mean) for all types	$t_{amb}=25°C$	$t_{amb}>25°C$ *Reduction/°C*
(1) Clamped directly on to 50 sq. in. of 16 S.W.G. aluminium	*10W*	*200mW*
(2) Clamped directly on to 9 sq. in. of 16 S.W.G. aluminium	*4W*	*80mW*
(3) As (2) but with 2 mil mica washer between heat sink and transistor	*2W*	*40mW*
(4) Transistor only in free air	*1W*	*20mW*

* **High power rating—up to 10W at audio and supersonic frequencies.**
* **High current ratings up to 3A DC.**
* **Long life.**
* **Excellent resistance to mechanical shock.**
* **Hermetic sealing and rigorous manufacturing control ensure uniformity and stability of a high order.**

British Design, Materials and Craftsmanship

Data sheets gladly forwarded on request

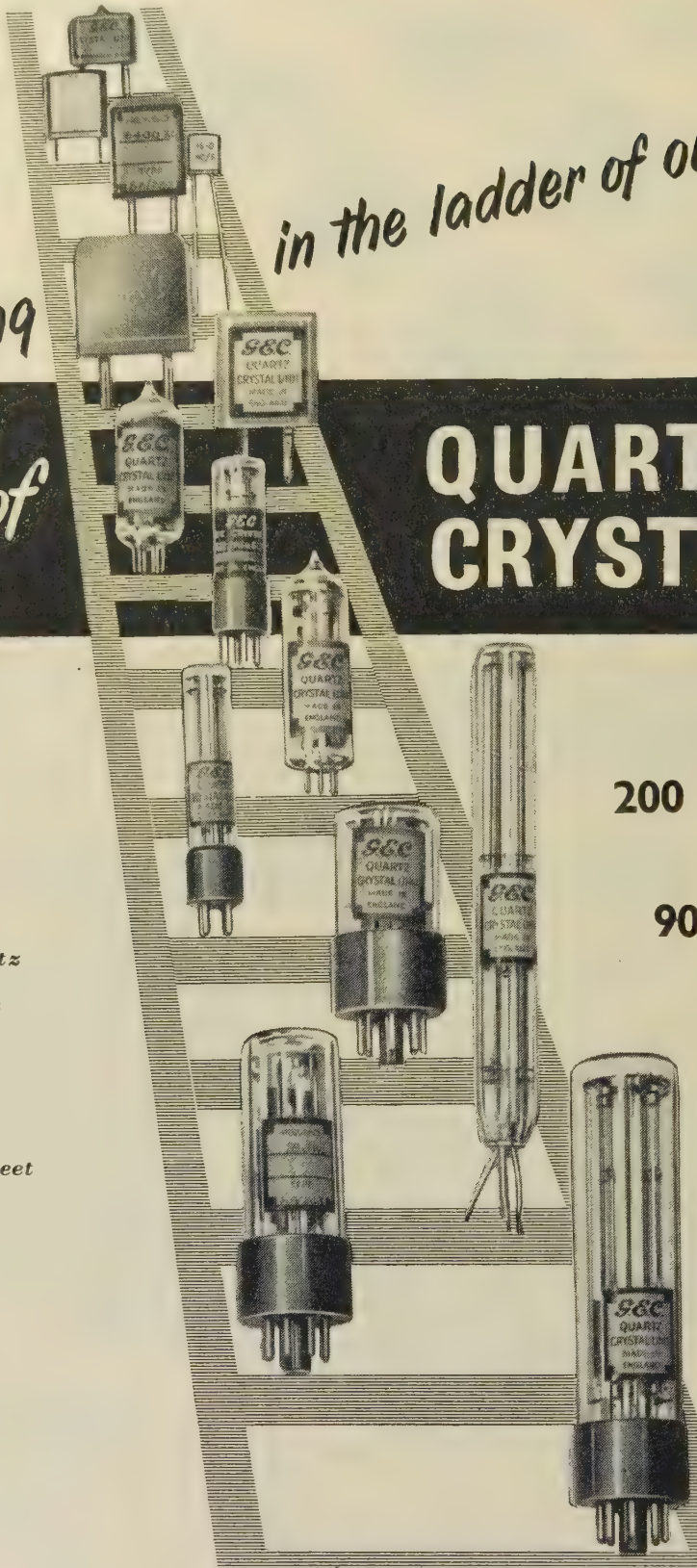
All trade enquiries to: **Newmarket Transistor Co. Ltd.**

Exning Road, Newmarket. Telephone: Newmarket 3381/4

TA 10705

no rungs missing in the ladder of our range

of **QUARTZ CRYSTALS**



G.E.C.

From 200 cycles/sec. to 90 Mc/sec.

For long term stability and unfailing activity, G.E.C. Quartz Crystal Units provide the basis for reliable communications systems.

A complete range of units to meet D.E.F.5271 and R.C.L.271 Inter-Services styles can be supplied.

SALFORD ELECTRICAL INSTRUMENTS LIMITED

(COMPONENTS GROUP)

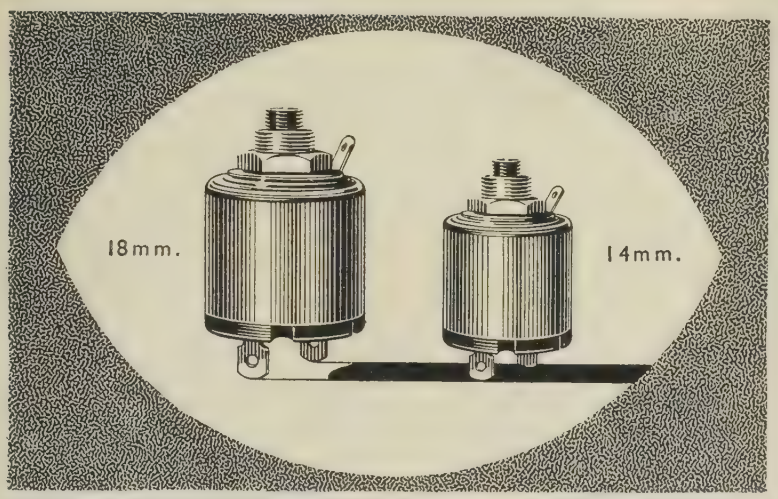
TIMES MILL · HEYWOOD · LANCASHIRE Tel: Heywood 6868

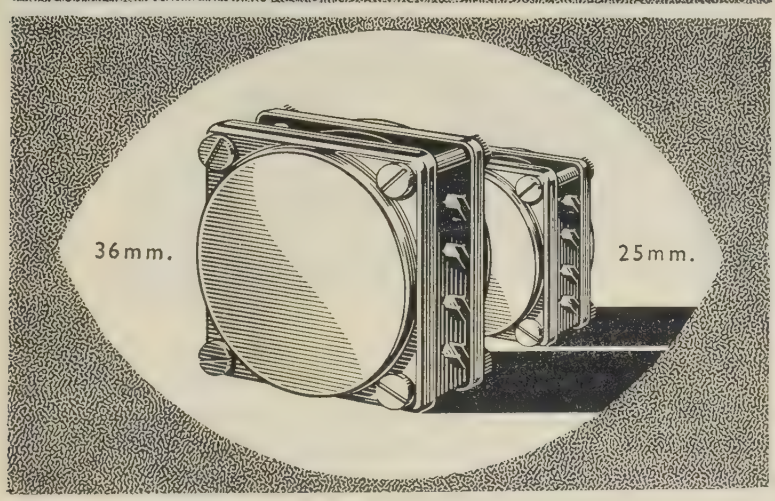
London Sales Office Tel: Temple Bar 4669

A SUBSIDIARY OF THE GENERAL ELECTRIC CO. LTD. OF ENGLAND

MULLARD HIGH EFFICIENCY POT CORES

have these outstanding features .....





* Pot core design facilitating rapid assembly
* Small size
* High value inductance
* Low losses resulting in high Q values
* Very fine setting accuracies
* Operative over a wide frequency range
* Controlled temperature coefficient

Wherever high quality pot cores are required, there will be a Mullard type available to meet the specification, furthermore, they can be supplied wound to customers individual requirements.

Write now for full details of the comprehensive range currently available.

Mullard

'Ticonal' permanent magnets
Magnadur ceramic magnets
Ferroxcube magnetic cores

MULLARD LTD., COMPONENT DIVISION, MULLARD HOUSE, TORRINGTON PLACE, W.C.1

MC 255A

Accident or Design?

Shown by courtesy of K.L.G. SPARKING PLUGS LTD., for whom all the above components were made.

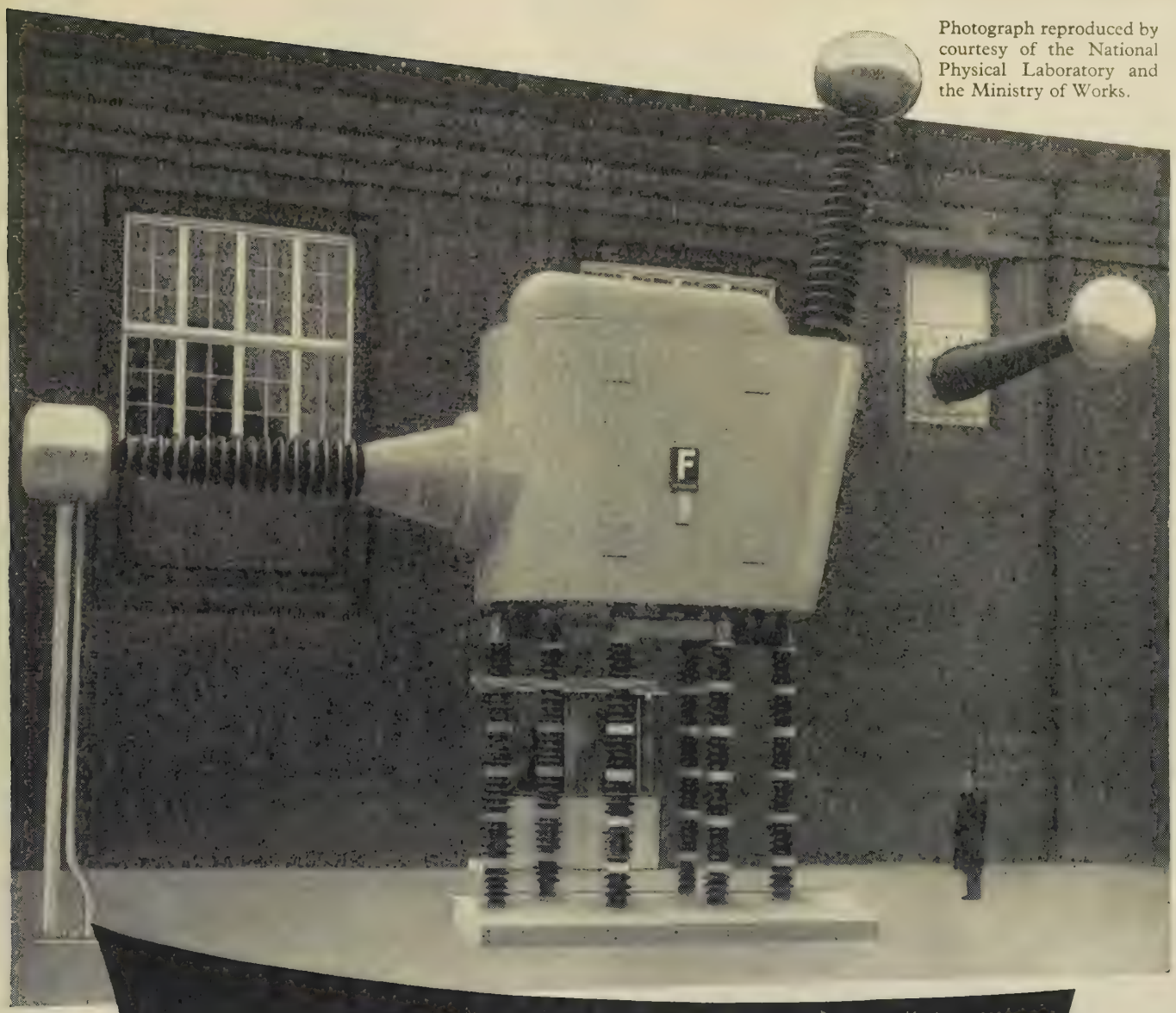
IT IS no accident that well-known progressive concerns, year after year, turn to us for bulk supplies — running into millions of components — of high quality, reasonably priced, Plastic mouldings — delivered on, *or before*, time.

You, too, will do the same when you know us.

CERTIFIED ZINC ALLOY ZAMCA DIE CASTERS

METROPOLITAN PLASTICS LTD

GLENVILLE GROVE · DEPTFORD · LONDON SE8 *Phone* TIDeway 1172-3



Photograph reproduced by courtesy of the National Physical Laboratory and the Ministry of Works.

1,000,000 VOLT FERRANTI TESTING TRANSFORMER *installed at the* NATIONAL PHYSICAL LABORATORY

THE HIGHEST VOLTAGE OUTDOOR TESTING TRANSFORMER IN THE UNITED KINGDOM

A 1,000,000 Volt Ferranti Testing Transformer operating with the tank at a potential of 500,000 Volts above earth. The transformer is mounted on 6 columns of insulators at a height of 12 feet from ground level. The overall height of the complete equipment is 36 feet. Associated with the equipment is a 1,000,000 Volt wall bushing, 36 feet in length, through which the supply is carried to the test laboratories.

Ferranti are specialists in the manufacture of High Voltage A.C., D.C. and Impulse Testing Equipments for all kinds of research and routine testing and are backed by over 70 years' experience in the manufacture and development of electrical apparatus.

This equipment was installed under the direction of the Ministry of Works.

FERRANTI LTD • HOLLINWOOD • LANCS

London Office : KERN HOUSE · 36 KINGSWAY · W.C.2.

FT 201

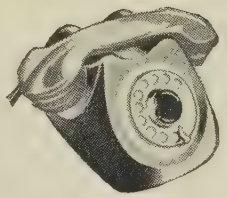
SIEMENS EDISWAN introduce

the first print

SIEMENS EDISWAN

SIEMENS EDISON SWAN

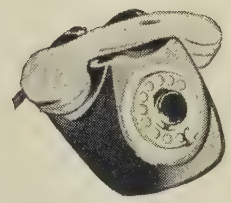
An AEI Company Woolwich, London, S.E.18.



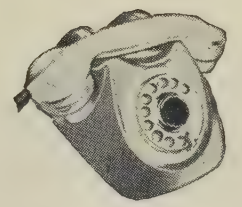
THE CENTENARY NEOPHONE

A quality instrument at an economical price, especially suitable for use overseas.

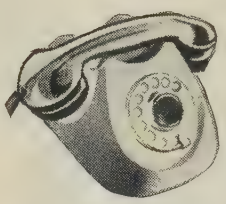
circuit telephone



This entirely new instrument, a product of the designers of the world-famous "Neophone" which established new standards in telephone performance and appearance in the nineteen-thirties, incorporates the following attractive features —



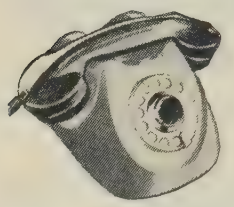
Improved performance, using latest design of components.
Lower first cost, achieved by the most modern manufacturing methods
Lower maintenance cost.
Suitably finished for tropical use. Sealed case keeps out dust and insects.
Reduced size and weight – handset weight halved – only 7 ounces.



21 **COLOUR COMBINATIONS**

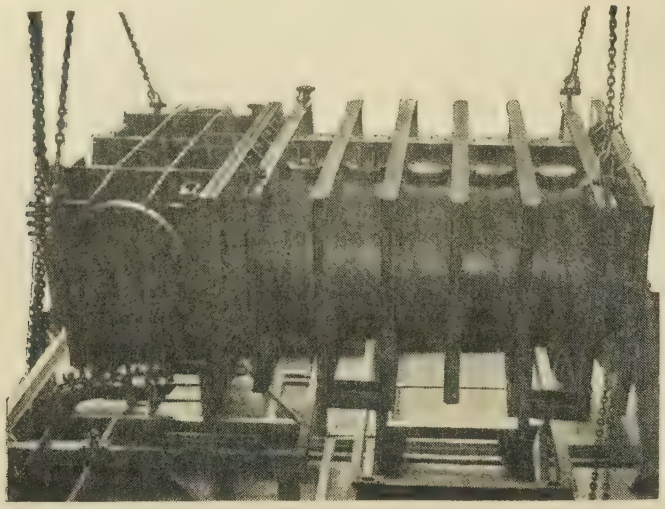
7 different handset shades and a choice of 3 case colours. The case can be changed WITHOUT DISTURBING THE DIAL SWITCH.

The centenary NEOPHONE



THE NEW INSTRUMENT WITH NEW FEATURES

Complete technical details are available on request.



a cyclotron for medical research

Robert Jenkins & Co. Ltd., of Rotherham, are proud to have supplied the vacuum vessel and the vacuum pumping manifold for the Cyclotron purchased by the Medical Research Council for radio-therapeutic research.

The first photograph shows the commencement of the assembly with the two components being lowered on to the supporting trolley.

In the second photograph you see a parallel section of the vacuum vessel and the vacuum pumping manifold in position with the complete assembly.

Robert Jenkins & Co. Ltd.
ESTABLISHED 1856
ROTHERHAM

IVANHOE TRADE MARK
THE SIGN OF GOOD WELDING

Telephone: 4201–6 (6 lines)

FOR THE M.R.C. CYCLOTRON

A pair of Dees in Copper were made to the requirements of the Medical Research Council for 45″ Cyclotron at Hammersmith Hospital, London.

We have previously supplied the Dees for the Cyclotrons at A.E.R.E., Harwell, and Birmingham University.

We should be pleased to receive enquiries for fabrications in copper to customers' special requirements.

LEE & WILKES, LTD.

Tube Manipulators - Coppersmiths - Metal Spinners

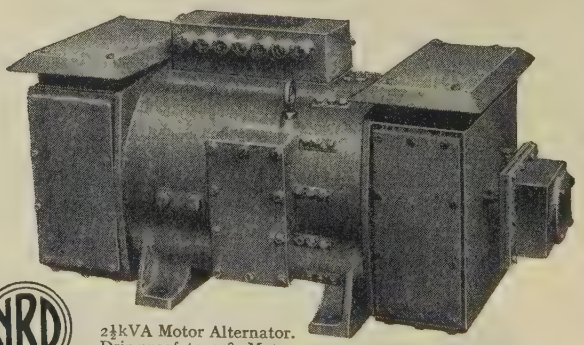
PRIORY COPPER WORKS, BREWERY STREET, BIRMINGHAM 6
Phone: Aston Cross 2005-6-7 Grams: "Kettle Phone Birmingham"

NEWTON-DERBY
ELECTRICAL EQUIPMENT

High Frequency Alternators

(Send for Publication No. 1003/2)

Also makers of Rotary Transformers and Converters, Wind and Engine-Driven Aircraft Generators, High Tension D.C. Generators, and Automatic Carbon Pile Voltage Regulators.

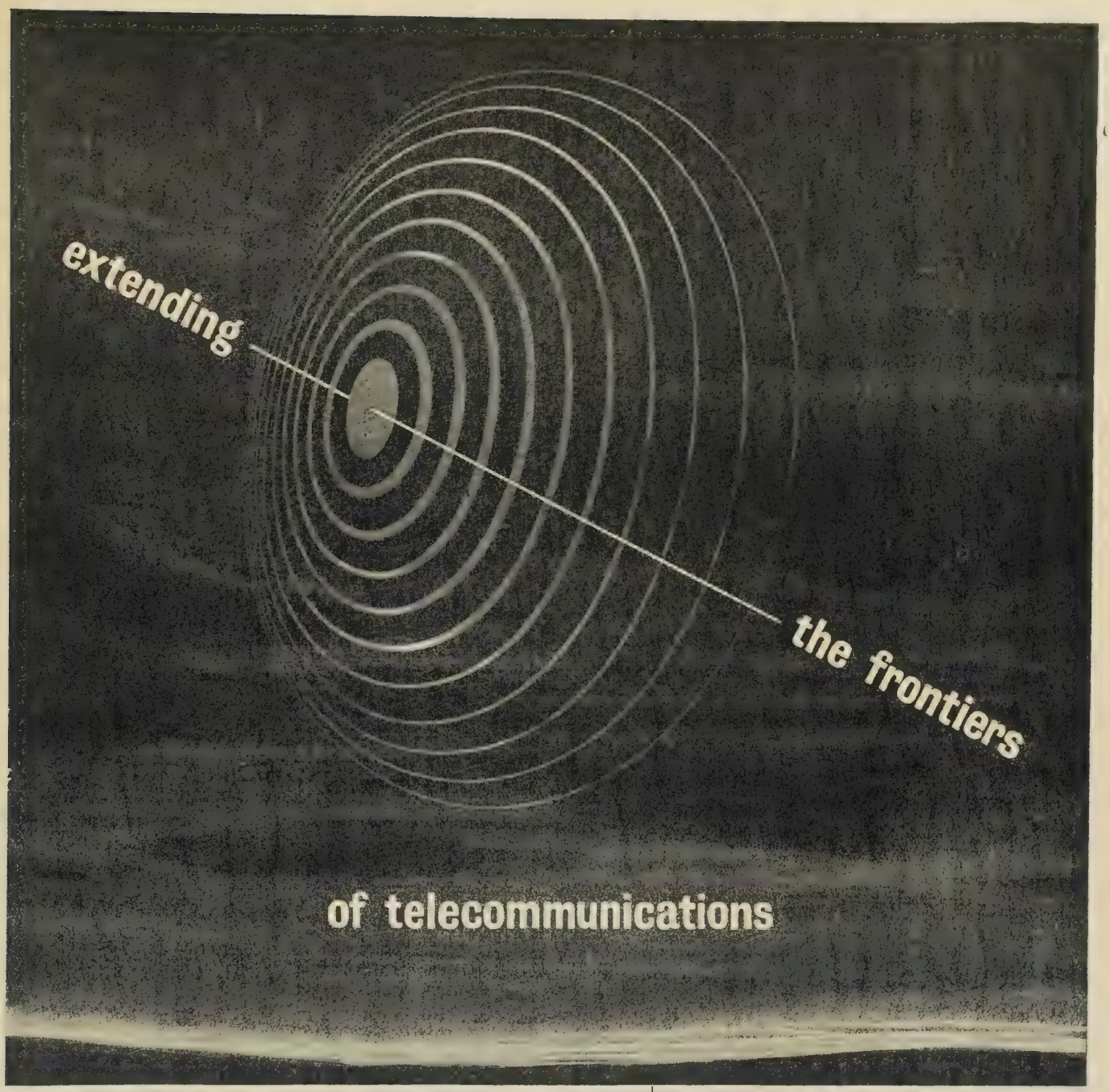


NBD

2½kVA Motor Alternator. Drip proof to 45°. Motor 220 volts D.C. Output 120 volts. 3 phase. 333 cycles per second. Motor includes an automatic constant speed governor. Weight 450 lb.

NEWTON BROTHERS
(DERBY) LTD

HEAD OFFICE & WORKS: ALFRETON ROAD, DERBY
TELEPHONE: DERBY 47676 (4 lines) TELEGRAMS: DYNAMO, DERBY
LONDON OFFICE: IMPERIAL BUILDINGS, 56 KINGSWAY W.C.2



SIEMENS EDISWAN is rapidly becoming an important name in the world of telecommunications. Faithful adherence to the high standards of manufacture and testing established for so many years by Siemens Brothers and Ediswan, coupled with an extensive programme of continuous research and design, are the keystones of this progress. They are the reasons why Siemens Ediswan are gaining a fine reputation as makers of telecommunications equipment second to none in dependability and efficiency.

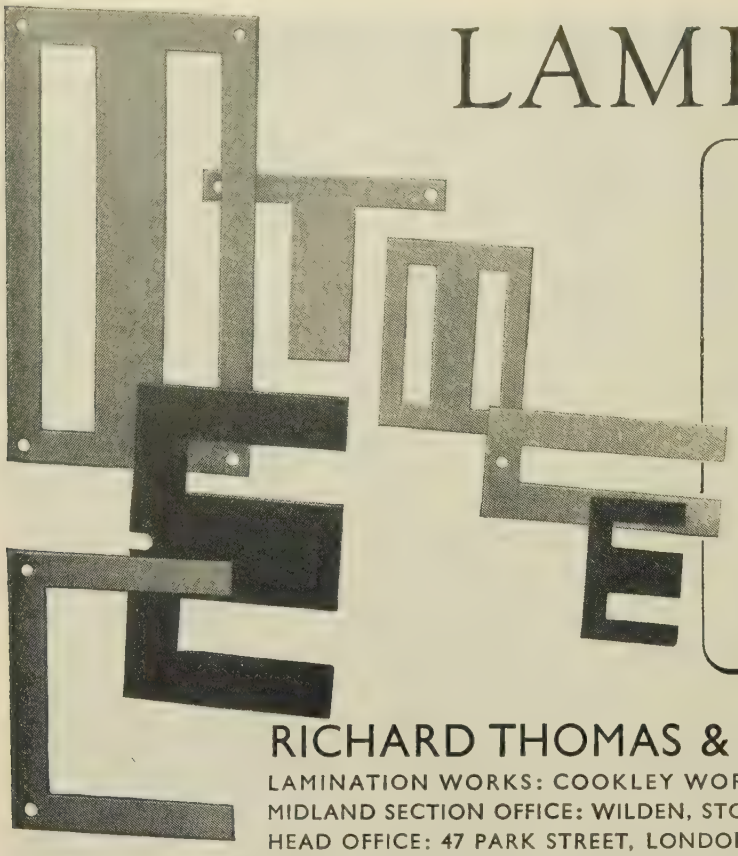
Details of our current range of equipment are available on application.

BROAD BAND CARRIER TELEPHONE EQUIPMENT
PROGRAMME CHANNEL EQUIPMENT
VOICE FREQUENCY TELEGRAPH EQUIPMENT
VOICE FREQUENCY REPEATER EQUIPMENT
RADIO TELEPHONE TERMINAL EQUIPMENT
FIVE BAND PRIVACY EQUIPMENT
TRANSMISSION TEST GEAR
COMPONENTS AND OTHER EQUIPMENT

SIEMENS EDISWAN

SIEMENS EDISON SWAN LTD *An A.E.I. Company*
Telecommunications Transmission Division, Woolwich, London SE18
Telephone: Woolwich 2020 Telegrams: Siemswan Southe London

11/64



LAMINATIONS

Laminations of all types, in all sizes and in all grades of material

FERROSIL
hot-rolled and cold-reduced electrical sheet and strip, and hot-rolled transformer sheet

ALPHASIL
cold-reduced oriented transformer sheet and strip

RTB

RICHARD THOMAS & BALDWINS LTD
LAMINATION WORKS: COOKLEY WORKS, BRIERLEY HILL, STAFFS.
MIDLAND SECTION OFFICE: WILDEN, STOURPORT-ON-SEVERN, WORCS.
HEAD OFFICE: 47 PARK STREET, LONDON, W.1

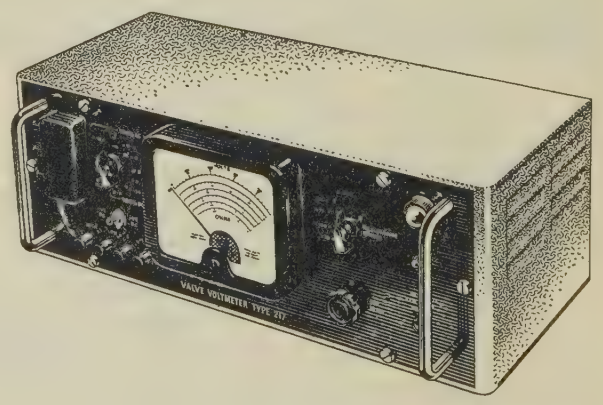
Our Cookley Works is one of the largest in Europe specializing in the manufacture of laminations for the electrical industry.

AIRMEC

VALVE VOLTMETER TYPE 217

THE Valve Voltmeter Type 217 is a redesigned version of the well-known Airmec Valve Voltmeter Type 712. An exceptional degree of stability has been obtained by the use of modern valves and improved probe mounting. The frequency response has also been extended whilst all the desirable features of its predecessor have been retained.

- **Frequency range from 20 c/s to 200 Mc/s**
- **Balanced, unbalanced and differential inputs**
- **Measures both positive and negative D.C. voltages**
- **Six resistance ranges reading up to 100 Megohms**
- **Balanced circuitry ensures exceptional stability**
- **Very low input capacity**



IMMEDIATE DELIVERY

Full details of this or any other Airmec instrument will be forwarded gladly on request.

AIRMEC LIMITED
HIGH WYCOMBE BUCKS ENGLAND
Telephone: High Wycombe 2060
Cables: Airmec High Wycombe

a.c./d.c. conversion

with Germanium Power Rectifiers

For all industrial d.c. power supplies, we offer the benefit of thirty years experience as Metal Rectifier Manufacturers.

Following copper oxide and selenium rectifiers, germanium rectifier equipments can now be provided for outputs above 25 kW where the superior efficiency and reduced bulk offers advantages.

We shall be pleased to supply details of germanium rectifier equipments.

Westinghouse Brake & Signal Co. Ltd.

(Dept. I.E.E. 9)
82 York Way, King's Cross, London, N.1.
Tel. TERminus 6432

High-Stability Wire Wound Precision Resistors

Felgate Electronics Limited announce a new range of Wire Wound Precision Resistors which offer utmost reliability, prompt delivery, and exceptionally high stability (0·02 %). The specification speaks for itself:

THE STANDARD RANGE 0·1Ω to 4MΩ resistors outside this range can be manufactured to special order.

Accuracy up to ± 0·05 % or 0·01Ω whichever is greater. Matched pairs can be supplied to even greater accuracy.

Temperature co-efficient of resistance (α) — Two values both guaranteed:
Cu/Ni resistance wire $\alpha = 0{\cdot}002\%$ per °C
Ni/Cr resistance wire $\alpha = 0{\cdot}01\%$ per °C

Resistors can be manufactured to attain the required value of resistance at such ambient temperatures and loading as are specified by the customer.

Resistors from ¼ watt to 2 watts are available. All types are back-to-back wound for minimum self-inductance.

TROPICALISATION All Felgate Resistors are encapsulated in a robust resin to be proof against humidity.

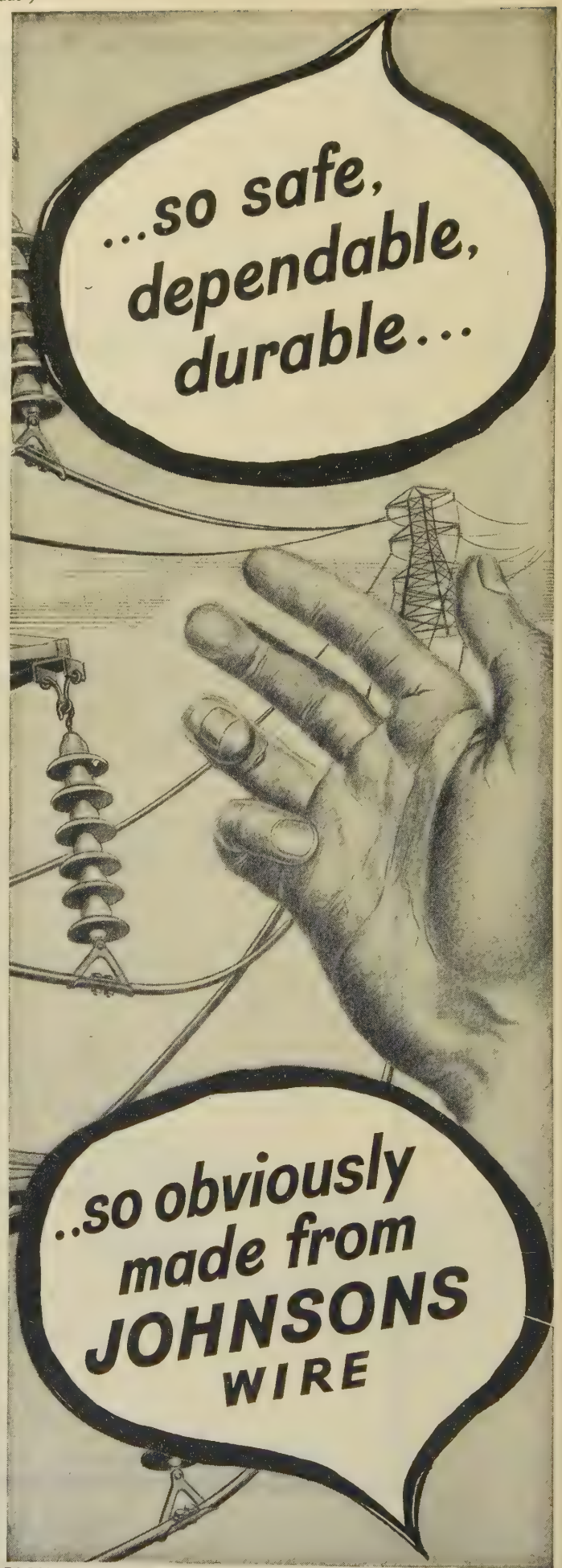
SPECIAL TYPES are available including American equivalents.

DELIVERY Owing to extremely versatile manufacturing techniques, prompt deliveries can be made against special orders, whether for home or export markets.

RELIABILITY is recognised as being of cardinal importance to the manufacturers of electronic equipment—Felgate components are designed with this in mind.

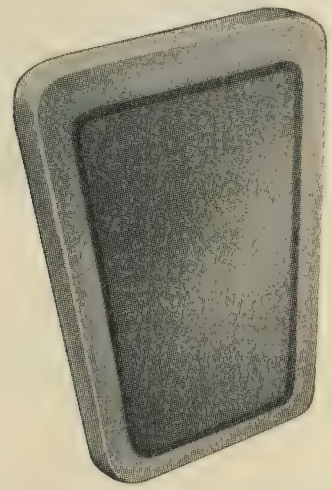
For further details please write to:

FELGATE ELECTRONICS LIMITED
Felgate House, Studland Street, Hammersmith, W.6
Tel. Riverside 8141/2.



RICHARD JOHNSON & NEPHEW LIMITED, MANCHESTER 1.

The SPADE Ceramicon*



THE Capacitor for Printed Circuits

Fixing is ridiculously simple. Cut a dog bone hole like this , push in the , dip solder, and the job is done.

- **No leads to bother with.**
- **No expensive insertion equipment.**
- **Maximum economy in space.**
- **Speed in assembly.**
- **Minimum inductance.**

SPECIFICATION

- **Dielectrics: Temperature compensating and high permittivity.**
- **Capacitance values: 2 pF to 1,000 pF.**
- **Working volts: 500 D.C.**

Supplied in sealed polythene bags with anti-oxidant.

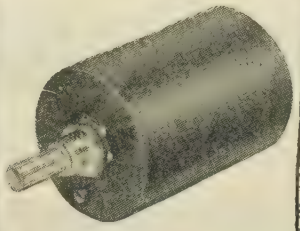
ERIE* *Resistor Ltd*

★ Registered Trade Marks

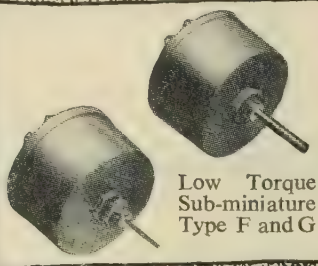
Carlisle Road, The Hyde, London, N.W.9., England. Tel: COLindale 8011. *Factories: London and Great Yarmouth, England; Toronto, Canada; Erie, Pa., and Holly Springs, Miss., U.S.A.*

Precision Potentiometers . . .

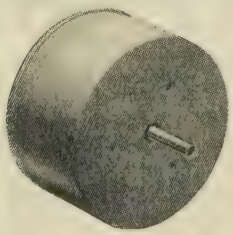
by P·X·F



Helical Potentiometer Type PX 4/H.10. Turnings: 3 and 10.



Low Torque Sub-miniature Type F and G



Toroidal Potentiometer Type B and E with ball races.

Ceramic Potentiometer 8 sizes 10–1,000 watts.

You want reliability? Then you want Fox's! Our potentiometers are renowned throughout the world for their simplicity, amazing mechanical strength, and, most important of all, reliability under the most vital and critical conditions.

May we advise you on any problem you may have? Please write for details of the large range of toroidal and helical potentiometers.

P. X. FOX LIMITED

Specialist manufacturers of Toroidal Potentiometers
Dept. 215.
Hawksworth Road, Horsforth, Yorkshire.
Tel.: Horsforth 2831/2.
Grams.: "Toroidal, Leeds."

ndh 887

THE JOURNAL OF

The British Nuclear Energy Conference

The Institution of Civil Engineers · The Institution of Mechanical Engineers
The Institution of Electrical Engineers · The Institute of Physics
The Institution of Chemical Engineers

PUBLISHED JANUARY, APRIL, JULY, OCTOBER

The Journal contains papers and discussions on the applications of nuclear energy and ancillary subjects

ANNUAL SUBSCRIPTIONS:

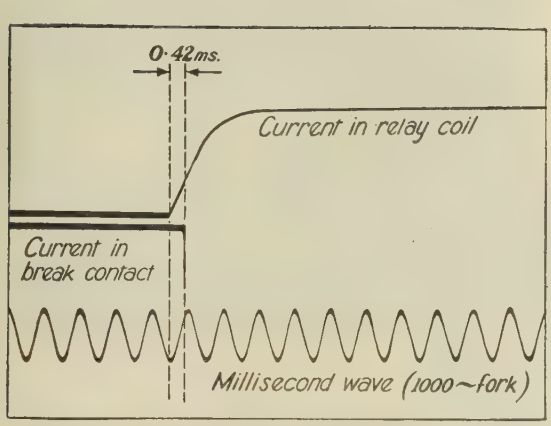
MEMBERS 30/- post free
NON-MEMBERS 60/- post free

Full particulars are available from

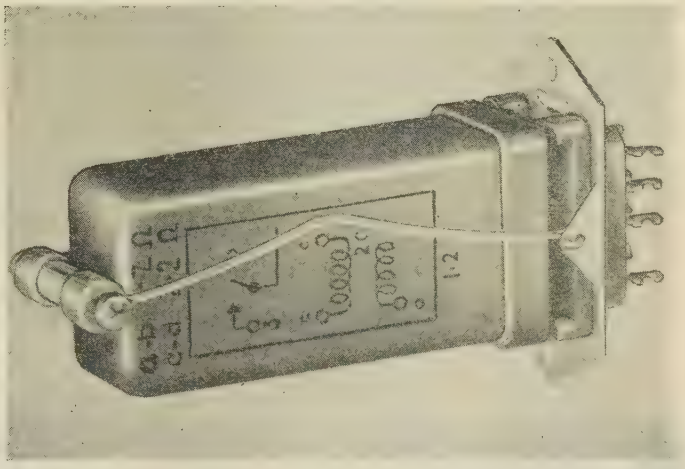
The Secretary • B.N.E.C. • 1-7 Great George Street • London • SW1

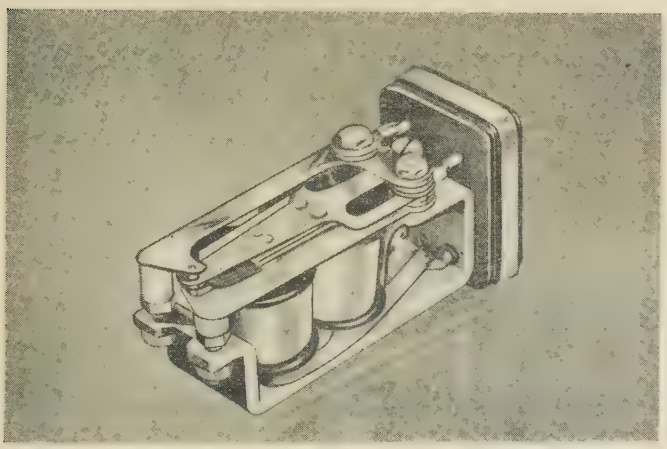
SIEMENS EDISWAN

Miniature *high speed* relay



Oscillogram of the operating-lag of the break contact





This small relay has earned for itself a world-wide reputation on account of its very rapid operation, great reliability and insensitivity to external mechanical and electrical disturbances. It was originally designed to meet very severe conditions called for by the Services; it is now in use in vast numbers for all manner of applications and varying conditions.

Inexpensive and available for early delivery.

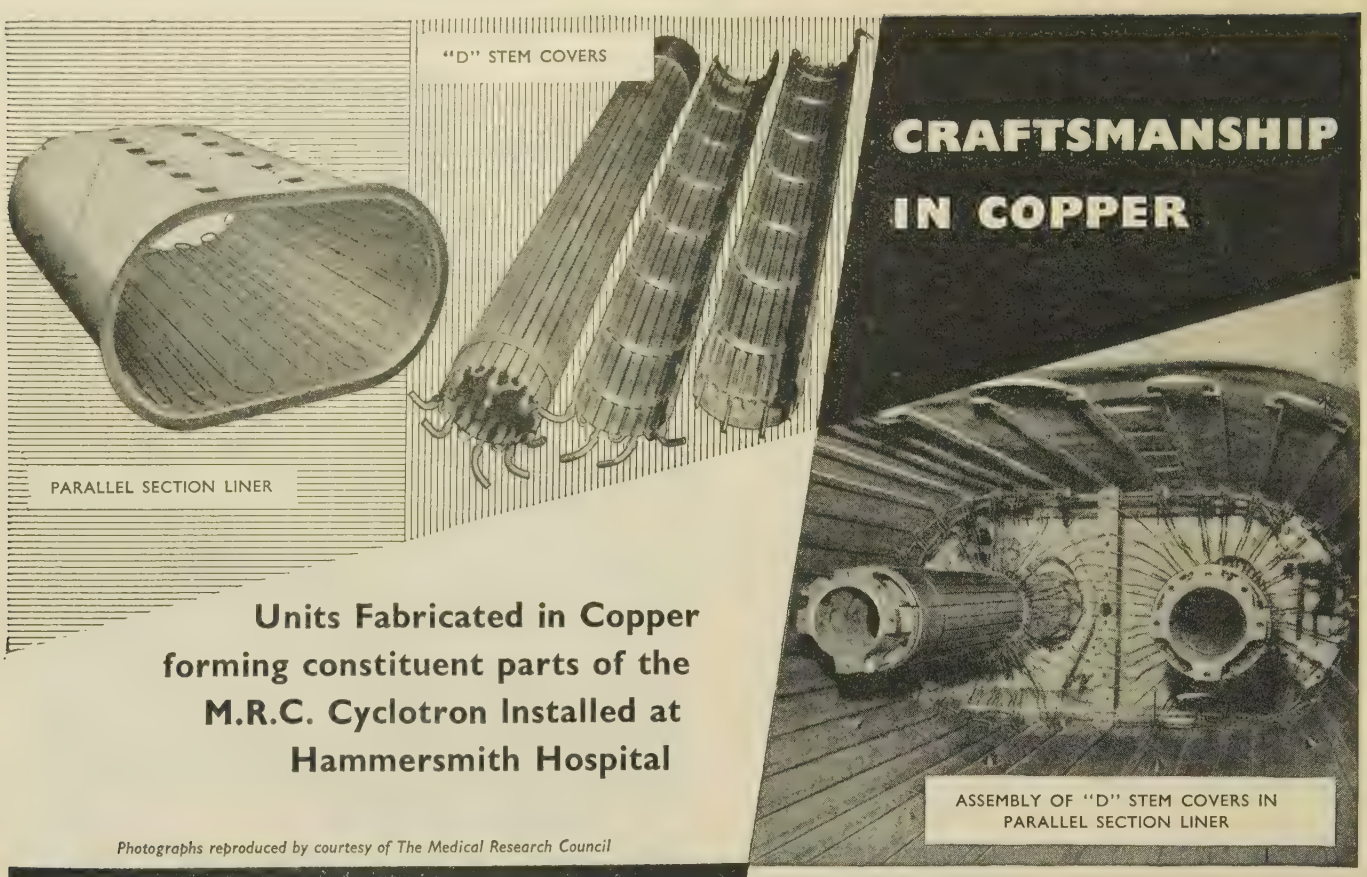
Can be supplied with plug-in base if required to form a readily interchangeable plug-in unit.

Hermetically sealed, unaffected by dirt or immersion in water and immune to wide or rapid changes of temperature or air pressure.

We shall be pleased to give full details if you telephone (Woolwich 2020, Extension 621) or write to:

SIEMENS EDISON SWAN LTD *An A.E.I. Company*

WOOLWICH LONDON SE18



CRAFTSMANSHIP IN COPPER

Units Fabricated in Copper forming constituent parts of the M.R.C. Cyclotron Installed at Hammersmith Hospital

ASSEMBLY OF "D" STEM COVERS IN PARALLEL SECTION LINER

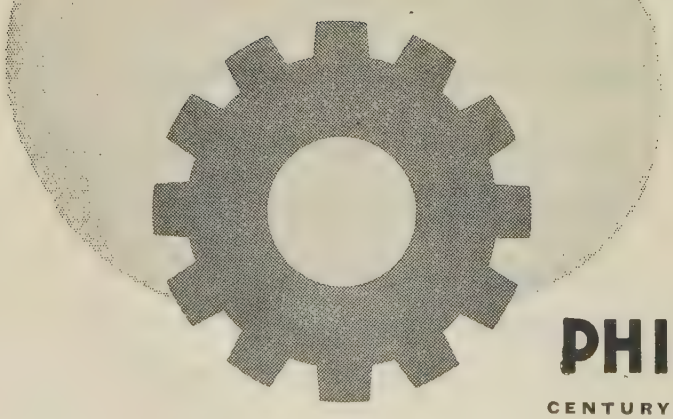
Photographs reproduced by courtesy of The Medical Research Council

CUPROCYL LTD.

Cuprocyl Works:
230 YORK WAY, LONDON, N.7.
Phone: North 4887/8 · Grams: CUPROCYL, HOLWAY, LONDON

PHILIPS

In Science and Industry alike . . .



among technicians, manufacturers and those engaged in the sale of electrical products — as well as among the public at large, the Philips emblem is accepted throughout the World as a symbol of quality and dependability.

PHILIPS ELECTRICAL LTD

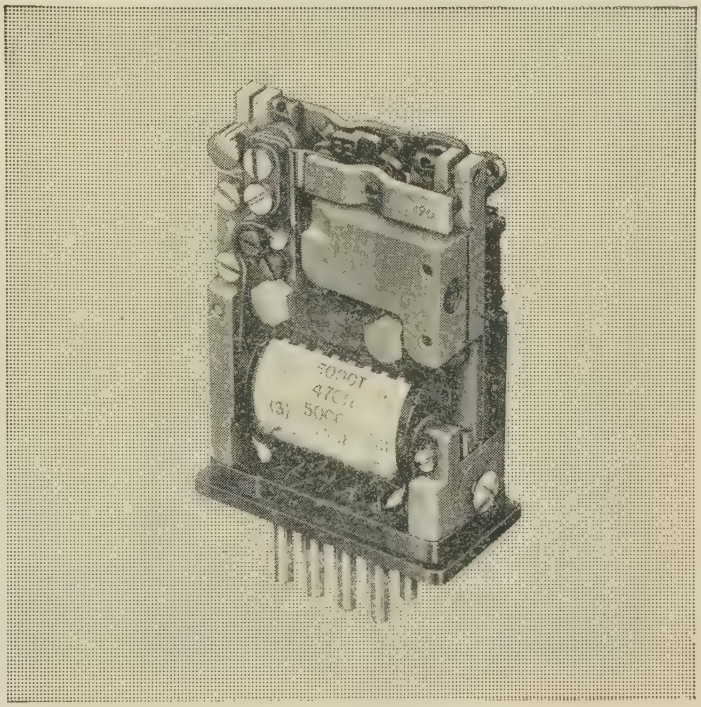
CENTURY HOUSE, SHAFTESBURY AVENUE, LONDON, W.C.2

RADIO & TELEVISION RECEIVERS · RADIOGRAMS & RECORD PLAYERS · GRAMOPHONE RECORDS · TUNGSTEN, FLUORESCENT, BLENDED AND DISCHARGE LAMPS & LIGHTING EQUIPMENT · 'PHILISHAVE' ELECTRIC DRY SHAVERS · 'PHOTOFLUX' FLASHBULBS · HIGH FREQUENCY HEATING GENERATORS · X-RAY EQUIPMENT FOR ALL PURPOSES · ELECTRO-MEDICAL APPARATUS · HEAT THERAPY APPARATUS · ARC & RESISTANCE WELDING PLANT AND ELECTRODES · ELECTRONIC MEASURING INSTRUMENTS · MAGNETIC FILTERS · BATTERY CHARGERS AND RECTIFIERS · SOUND AMPLIFYING INSTALLATIONS · CINEMA PROJECTORS · TAPE RECORDERS (P23)

Now in production!

The Centre-Stable

TYPE 51M CARPENTER POLARIZED RELAY



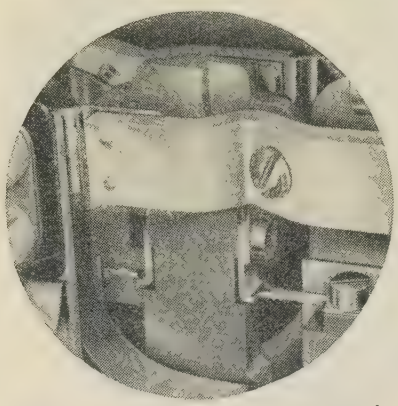
DIMENSIONS (excluding connecting tags and guide pins): HEIGHT: $2\frac{3}{8}$in. WIDTH: $1\frac{5}{32}$in. DEPTH: $\frac{25}{32}$ in.

This new highly sensitive *centre-stable** polarized relay is a development of the well-known Type 5 Carpenter Relay.

Like its companion—the *each-side-stable* Type 51A relay—the 51M's high performance is due largely to a new form of armature suspension, and the provision of variable slugs for adjusting the flux of the polarizing magnets thereby enabling the magnetic and mechanical parameters to be balanced with greater precision than has hitherto been possible.

With a sensitivity of approximately 10μW, the Type 51M relay is ideally suited for use as a sensing relay in Servo Systems, or as a detector in self balancing bridge circuits. Fitted with S.P.D.T. contact action, the relay is available in sealed or non-sealed forms.

For further details send for leaflet F.3526 which gives information of operating characteristics and the available range of coils.



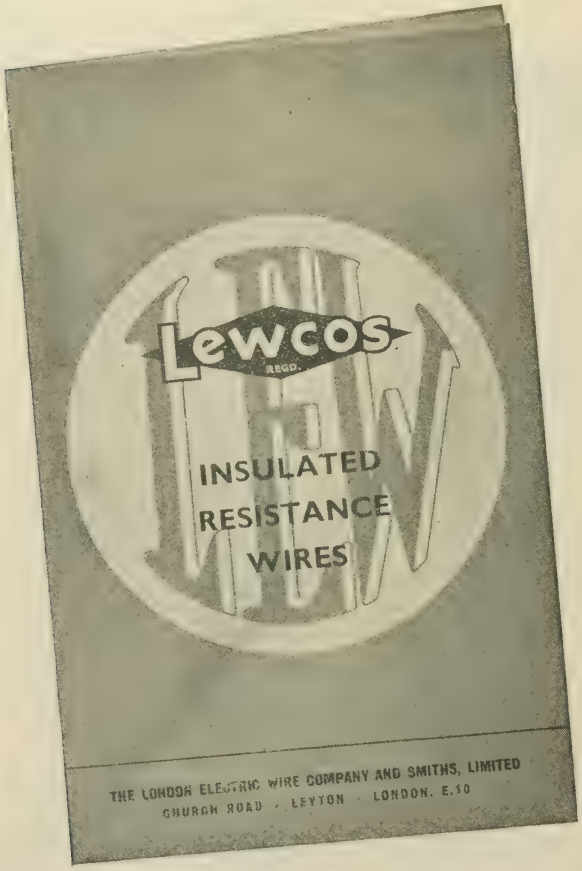
Close-up of armature suspension —magnet removed

** In the centre-stable form, the armature of the relay remains in a central position between the side-contacts when the relay is unenergized, and moves to one or other side-contact when current is applied to the coil.*

Manufactured by the sole licensees:

TELEPHONE MANUFACTURING COMPANY LTD
HOLLINGSWORTH WORKS, DULWICH, LONDON SE21

Telephone: GIPsy Hill 2211



Fine wires always in demand for precision work

Lewcos Insulated Resistance Wires with standard coverings of cotton, silk, glass, asbestos, standard enamel and synthetic enamel are supplied over a large range of sizes.

Send for our new leaflet

THE LONDON ELECTRIC WIRE COMPANY AND SMITHS LIMITED

LEYTON LONDON E.10

ADCOLA
PRODUCTS LIMITED
(Regd. Trade Mark)

SOLDERING INSTRUMENTS & ALL ALLIED EQUIPMENT

BIT SIZES
⅛" to a ½"

VOLT RANGES FROM 6/7 to 230/50 VOLTS
WITH NO EXTRA COST FOR LOW VOLTAGES

ASSURE *SOUND* JOINTS FOR *SOUND* EQUIPMENT

ADCOLA
PRODUCTS LTD.
Head Office & Sales
GAUDEN ROAD
CLAPHAM HIGH St.
LONDON, S.W.4

TELEPHONES
MACaulay 4272
MACaulay 3101

ZENITH
(REGD. TRADE-MARK)

PHASE SHIFTING TRANSFORMER

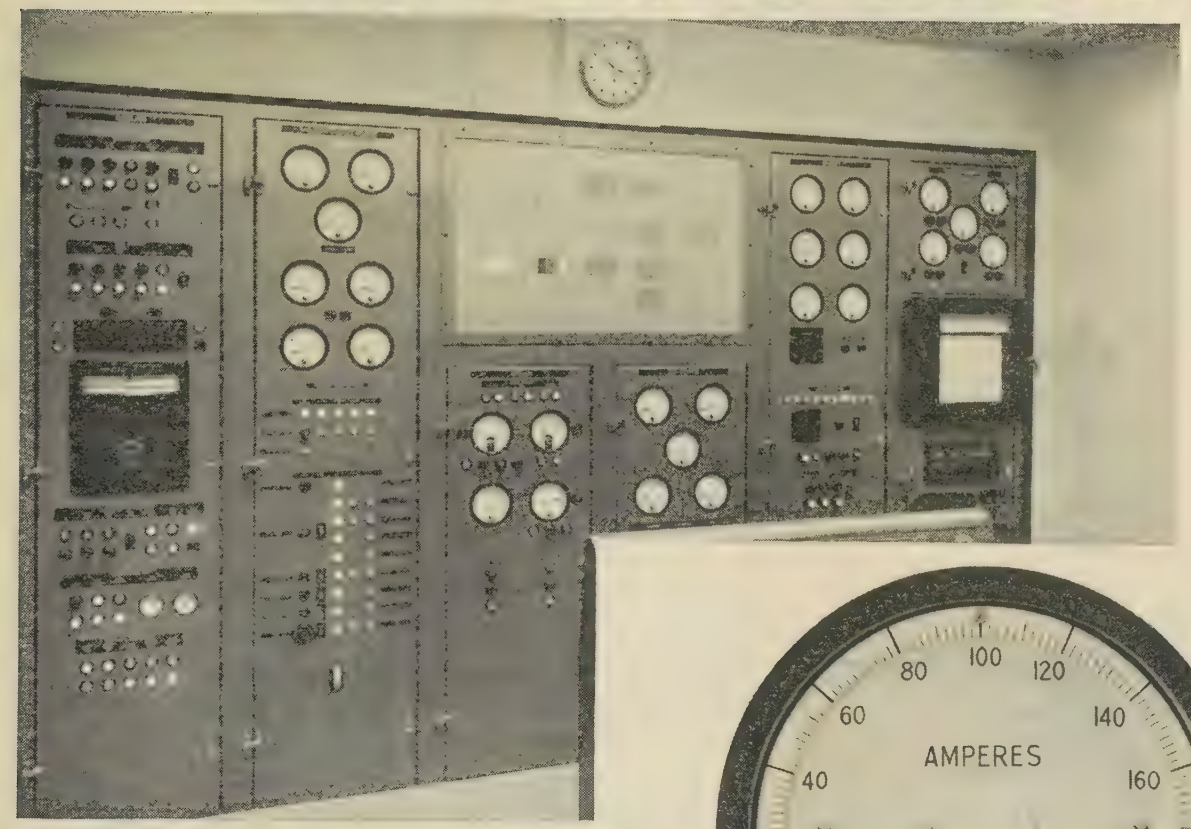
This instrument provides convenient means for adjusting the phase angle or power factor in alternating current circuits when testing single or polyphase service meters, wattmeters, or power factor indicators, etc. It is also the simplest means for teaching and demonstrating Alternating Current Theory as affecting phase angle and power factor.

Illustrated brochure free on request

The ZENITH ELECTRIC CO. Ltd.
ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREEN
LONDON, N.W.2

Telephone: WILlesden 6581-5 *Telegrams:* Voltaohm, Norphone, London
MANUFACTURERS OF ELECTRICAL ENGINEERING PRODUCTS INCLUDING RADIO AND TELEVISION COMPONENTS

NALDERS INSTRUMENTS



NALDERS INSTRUMENTS

on the control board for the M.R.C. CYCLOTRON
at the Hammersmith Medical Research Establishment

- HIGH SENSITIVITY AND ACCURACY
- CIRCULAR PLATFORM TYPE SCALE
- SPECIALLY CLEAR MARKINGS
- BAR TYPE POINTERS

We can meet your demands for
PROMPT DELIVERY

NALDERS PRODUCTS Include
MEASURING INSTRUMENTS of ALL TYPES
also
PROTECTIVE RELAYS
VECTORMETERS
Automatic Earth Proving SUPPLY SWITCHES
"BIJOU" CIRCUIT BREAKERS

NALDER BROS & THOMPSON LTD

DALSTON LANE WORKS, LONDON, E.8
Telephone: Clissold 2365 (3 lines)
Telegrams: "Occlude, Hack, London"

B.M.8

STANDARD TELEPHONES AND CABLES LIMITED · STANDARD TELEPHONES AND CABLES LIMITED · STANDARD TELEPHONES

S.T.C. have always led in pioneering telecommunications developments.

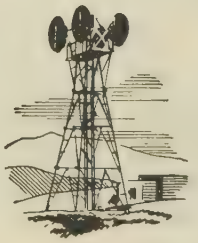
S.T.C. developed the first microwave system for 600 telephone channels.

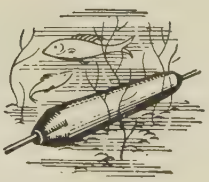
S.T.C. developed in 1946 the New Equipment Practice for the unit construction of slide-in panels.

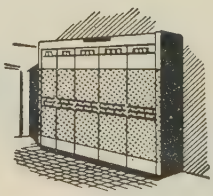
S.T.C. developed in 1951 a more compact form of unit construction which has been accepted as the model for up-to-date equipment.

S.T.C. developed the first major submerged repeater system in Europe.

S.T.C. developed in 1955 the new channel panel which has been adopted by leading organisations as the model for their present equipment.







The latest S.T.C. systems include:

Transistorised 12-channel, 2-wire, carrier-on-cable telephone system.

A transistorised 3-channel, open-wire, carrier telephone system.

A 7000 Mc/s portable television outside-broadcast system.

A transistorised 24-channel, frequency modulated, voice-frequency telegraph system.

A 12 Mc/s coaxial system handling up to 2700 high quality telephone circuits or 960 telephone circuits and a television channel.

A transistorised 10-channel, rural carrier system.

A small-diameter, coaxial cable telephone system for 240 channels using buried transistorised repeaters.

To be up-to-date follow S.T.C.

Leaders in Telecommunications Systems design

Standard Telephones and Cables Limited

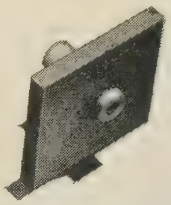
Registered Office: Connaught House, Aldwych, London, W.C.2

TRANSMISSION DIVISION: NORTH WOOLWICH · LONDON · E.16

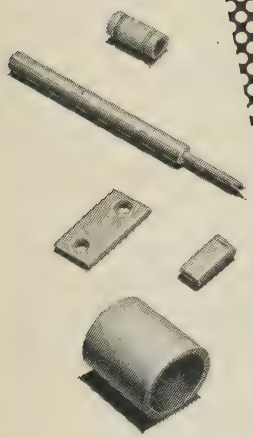
STANDARD TELEPHONES AND CABLES LIMITED · STANDARD TELEPHONES AND CABLES LIMITED · STANDARD TELEPHONES

INDEX OF ADVERTISERS

To illustrate the extreme hardness of SINTOX Industrial Ceramic, the photograph on the right shows a ½ in. diameter tube of SINTOX which, without even chipping, was forced through a ¼ in. thick mild steel plate.



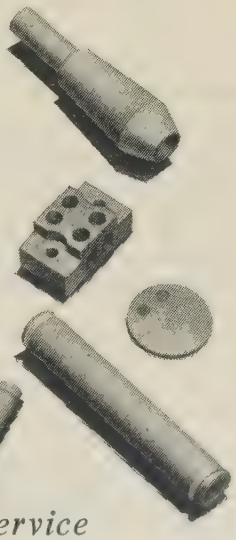
SINTOX IN THE ELECTRICAL INDUSTRY



Compared to porcelain, in tension, compression and cross breaking, SINTOX has at least twice the strength; and insulating properties twice as good.

Electrical resistance—SINTOX is so little affected by temperatures that at 300° C. it is about one million times that of porcelain. Thermal conductivity twenty times better than porcelain—almost comparable with steel.

SINTOX has already many applications in the electronics field. For instance, for formers in place of traditional wire-wound resistors and in the repeaters used for the Trans-Atlantic cable. Other applications in the electrical industry are legion.



THE HARD CERAMIC **SINTOX**

STAS

Sintox Technical Advisory Service

This service is freely available without obligation to those requiring technical advice on the application of Sintox Industrial Ceramics. Please write for booklet or any information required enclosing blue print if available.

SINTOX IS MANUFACTURED BY **LODGE PLUGS LTD., RUGBY**

'packaged' power

saves design time and effort

Performance figures unaffected by capacity loading on the H.T. output.

A COMPLETE RANGE OF SELF-CONTAINED UNITS READY TO BUILD INTO YOUR PROJECT	SRS 156	AS 619	AS 516	AS 517	AS 515	AS 615	AS 616
VOLTAGE	±150V	±150V to ±180V	±250V or ±300V	±250V or ±300V	±250V or ±300V	±250V or ±300V	±250V or ±300V
CURRENT	0-40ma	0-200ma	0-50ma	10-100ma	0-150ma or 50-200ma at 250V. 100-200ma at 300V.	0-500ma	0-1A
A.C. OUTPUTS	6.3V 4A C.T. 6.3V 1A	6.3V 4A C.T. 6.3V 4A C.T. 6.3V 1A C.T.	6.3V 4A C.T. 6.3V 1A	6.3V 4A C.T. 6.3V 2A C.T. 6.3V 1A	6.3V 4A C.T. 6.3V 4A C.T. 6.3V 1A	6.3V 5A C.T. 6.3V 5A C.T.	6.3V 10A C.T.
D.C. SOURCE IMPEDANCE	<2Ω	<1Ω	<1Ω	<1Ω	<1Ω	<1Ω	<1Ω
STABILISATION FACTOR BETTER THAN	400:1	400:1	400:1	400:1	300:1	400:1	400:1
A.C. SOURCE IMPEDANCE 40c/s—100kc/s	<2Ω	<2Ω	<0.25Ω	<0.25Ω	<0.5Ω	<0.5Ω	<0.5Ω
RIPPLE AND NOISE CONTENT	<350μV	<300μV	<250μV	<300μV	<1mV	<300μV	<300μV
DIMENSIONS	9⅛″×6½″ ×6¼″ high	10″×10″ ×6 9/16″ high	9⅛″×6½″ ×6¼″ high	9½″×7⅛″ ×6¼″ high	9⅜″×9⅛″ ×6⅜″ high	19″×12¼″ ×10½″ high	19″×19″ ×10½″ high
WEIGHT	14½ lbs	21¼ lbs	14½ lbs	20½ lbs	21¼ lbs	80 lbs	approx 150 lbs

From this complete range of regulated power supply sub-units you can choose, ready-made, the right power source for your prototype or production equipments. Why waste design time and effort on a part of the problem which we have already thoroughly covered for you?

These robustly made units are giving excellent service in many forms of electronic equipments, computers, simulators, industrial controls, etc. Educational and servicing establishments find them most useful as bench instruments because of their compactness.

Notable features are

- ★ LONG TERM RELIABILITY
- ★ LOW SOURCE IMPEDANCE
- ★ VERY LOW RIPPLE AND NOISE CONTENT
- ★ GOOD STABILISATION RATIO
- ★ LOW COST. DELIVERY EX STOCK

SOLARTRON

Power Supply Sub-units

THE SOLARTRON ELECTRONIC GROUP LTD • THAMES DITTON • SURREY • ENGLAND

Telephone: EMBerbrook 5522 • Cables: Solartron, Thames Ditton

The Institution is not, as a body, responsible for the opinions expressed by individual authors or speakers.
An example of the preferred form of bibliographical references will be found beneath the list of contents.

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 104. PART B. NO. 17. SEPTEMBER 1957

621.395.625.3

Paper No. 2332 R
Feb. 1957

THE 'STEREOSONIC' RECORDING AND REPRODUCING SYSTEM

A Two-Channel System for Domestic Tape Records.

By H. A. M. CLARK, B.Sc.(Eng.), Member, G. F. DUTTON, Ph.D., B.Sc.(Eng.), and P. B. VANDERLYN, Associate Members.

(*The paper was first received 7th September*, 1956, *and in revised form 4th January*, 1957. *It was published in February*, 1957, *and was read before the* RADIO AND TELECOMMUNICATION SECTION *20th February*, 1957.)

SUMMARY

The paper reviews briefly the history of stereophonic reproduction. The principal basic systems with their underlying ideas are described and compared. Some account is given of the supposed mechanism of natural binaural listening from the viewpoint of direction localization.

The principles and practice are discussed of a particular system for domestic use, derived from the early work of A. D. Blumlein, and characterized by the use of spaced loudspeakers driven in phase, to which the name 'stereosonic' has been given. The aims of this system are defined, and the mathematical theory involved in its use is developed. Limitations and sources of error in the results achieved are described.

Equipment used in the making of master recordings and some of the problems of studio technique involved are described. Consideration is given to the form which a domestic stereophonic record should take, and the standards to which such a record should conform, together with the requirements which these impose on the reproducing equipment.

LIST OF PRINCIPAL SYMBOLS

θ = Angle between radius vector of the sound source and the median plane.
θ_t = True angle.
θ_a = Apparent angle.
h = Effective human-ear spacing.
v = Velocity of sound in air.
ϕ = Phase difference of sound at the listener's two ears.
ω = Angular frequency.
ψ = Semi-angle subtended at the listener by a pair of loudspeakers.
μ = Difference of the time of arrival of the sound at the listener's two ears.
L and R = Peak amplitudes at listener's ears due to left and right loudspeakers.
x = Half distance between loudspeakers.
r = Projection of the listener's displacement from the median plane on to the loudspeaker base-line, expressed as a fraction of x.
d = Distance between loudspeaker base-line and the listener.
y_a = Fractional distance of apparent source from central position.
y_t = Fractional distance of true source from central position.
β and γ = Angles subtended by the left- and right-hand loudspeaker with respect to the median plane of the listener.
τ = Time-of-arrival difference of the sound from left- and right-hand loudspeakers at the listener caused by path differences.
μ_L and μ_R = Semi time differences at the listener's ears due to sounds from left and right loudspeakers.
ϕ_L and ϕ_R = Resultant phase of the sound at the left and right ears of the listener relative to the centre position.

(1) THE IMPORTANCE OF STEREOPHONY IN SOUND REPRODUCTION

Normal listening is always binaural, and the two ears are used in conjunction by the brain to interpret the sounds heard. If all the individual sound sources, such as voices and instruments, and the reverberant sounds, are recorded or transmitted by a single channel before being reproduced by one or more loudspeakers, it is not possible to use the binaural sense of location to differentiate between the various sounds as is done in direct listening.

In order to add some spatial realism using only one channel a number of schemes have been tried, such as the use of spaced loudspeakers supplied through frequency-dividing networks, or several loudspeakers fed through delay networks, but none of these is universally successful on all types of programme.

An honest assessment of the sound quality obtainable through a single channel, in which the frequency, harmonic, intermodulation and noise distortions have been reduced to the limits detectable by the human ear, reveals that the reproduction is still far from lifelike. The addition of real binaural listening conditions seems to be the most likely factor remaining.

The authors are with Electric and Musical Industries, Ltd.

Many tests have now been made which all show that, if a system is used in which the relative sounds at the two ears are each reproduced as similarly as possible to those heard in direct listening, i.e. with the correct relative amplitude and phase differences for each sound source, the reproduced sound takes on a new quality which is nearer to realism than can be obtained in any other way.

For a given amount of 'ambient' or reverberant sound, it is possible to obtain better definition, i.e. ability to distinguish individual instruments or voices in a chorus. A soloist can be heard clearly through an accompaniment, without having to resort to making the soloist appear unnaturally close to the listener.

The problem of hearing clear bass parts free from the unduly long low-frequency reverberation which is so commonly encountered, is considerably eased, and full bass response without 'boom' can be obtained. It is noticeable that the reproduced sound level can be raised higher than with single-channel reproduction before the average listener becomes irritated. The ability to do this permits a closer reproduction of the original dynamic range of the music in cases where this is great. It also seems possible to obtain a given subjective loudness for less acoustic power than with single-channel operation.

All this is quite aside from any benefits which may accrue from the ability to reproduce illusion of movements and the spatial relationships so essential to drama and opera.

The requirements then seem to be to reproduce, with reasonable economy, in a system suitable for home use, the same sounds at each ear separately which would have been heard by direct listening in an optimum position in front of the performers. This requirement can be limited to sounds arriving from directions covering an angle of, say, 90° in front of the observer for most practical purposes. No method has yet been devised to meet this requirement from all horizontal directions, but neither this nor the need for the indication of vertical angular direction seems to be of major significance.

(2) HISTORY

An explanation of the principles by which the normal human being uses both his ears for the purpose of spatial location has been sought since the classic studies in acoustics of the mid-nineteenth century. Attempts to reproduce this natural ability at a distance were made early in the development of telephony. One of the earliest accounts was given in *L'Électricien* of 1881 by E. Hospitalier, who described the sound system installed in the Paris Opera House.[1] This basically consisted of a pair of microphones spaced on opposite sides of the stage, which supplied at a distance a number of pairs of telephones held to the ears of the listeners, who received an impression of localization.

It was probably not for many years, until sound reproduction for entertainment purposes was receiving more attention, that any serious development work along these lines was undertaken. It was well understood, however, that the use of two similar microphones spaced by a distance equivalent to that between a human being's two ears, operating through two separate channels into a pair of headphones, did, in fact, give a remarkably accurate impression of spatial location. The use of headphones nevertheless robs the observer of the ability to locate sources outside the horizontal plane and prevents discrimination between front and rear, since in these cases it is necessary to move the head relative to the sound field.

In the 1920's it became customary to use loudspeakers for sound-reproducing systems instead of head telephones, and the problem of reproducing sound with directional sense immediately became more complex, for reasons which are discussed later in the paper.

One of the earliest workers in this field was the late A. D. Blumlein. He was convinced at this time that the main contribution towards the sense of spatial location was provided by the difference in time of arrival at the two ears which, at low frequencies, can be interpreted in terms of phase differences.

He invented a system[2] by which the phase difference between the outputs of two similar microphones spaced by a distance small compared with the wavelength, can be converted into two in-phase outputs of different amplitudes driving two loudspeakers spaced by an appreciable distance. If the relative amplitudes corresponding to a given phase difference at the microphones are correctly proportioned, the phase differences at the two ears of the person listening to the two loudspeakers will be a close simulation of the phase differences which the listener would have heard had he been at the microphone position.

Blumlein showed how the correct relative loudspeaker inputs could be derived from two closely spaced pressure microphones by the use of suitable modifying networks which he called 'shuffling' networks. He also showed how velocity microphones could similarly be used, and solved the mathematical relationship between the various parameters in order to obtain the correct apparent position of the source with any given loudspeaker arrangement. He also described a means of recording two channels on a single disc, and, in fact, disc records of this type were pressed and reproduced in 1933. Experimental motion-picture films were made incorporating twin-track optical recordings, and these were reproduced along with large-screen projection in 1935. The industry was not prepared to introduce the wide screen at this date, and therefore further work on the application was abandoned.

Blumlein's method has the advantage of being a 'free field' system, the observer being able to move his head without producing a corresponding movement of the apparent source. It relies on the interaction of sounds from the two loudspeakers at the listener's ears, and cannot utilize headphones.

One of the first successful public demonstrations of stereophonic sound reproduction was given[3] by Bell Laboratories in America in 1934. The system used three microphones at the centre and extremities, respectively, of the sound stage, and they drove, through three identical channels, three loudspeakers similarly disposed on the reproducing stage. Using such a three-channel system, fairly accurate location was obtained in the azimuth sense, and comparatively effective location was possible in depth. In 1939 R.C.A. used a similar arrangement with three channels for the stereophonic recording of a sound film entitled *Fantasia*.

Since the war demonstrations have been given by Philips at Eindhoven of a system using two microphones placed in an artificial head, the outputs of which supply two widely spaced loudspeakers. In such a system the phase differences of the microphone outputs become of no particular significance when reproduced at the two spaced loudspeakers, but the relative amplitudes of the microphone outputs due to masking by the artificial head are reproduced at the loudspeakers, and provide a measure of spatial location to an observer in front of them. Such an effect, of course, can only take place at the upper frequencies, for example those above about 700 c/s, and any directional effect at the low frequencies is neglected.

More recently several American companies have made recordings using two widely-spaced microphones driving spaced loudspeakers via the medium of twin-track magnetic tapes, and one has issued some discs carrying the two channels, one on the outer half and the other on the inner. This has the serious disadvantage of halving the playing time.

The advent of magnetic tape, providing two synchronous but otherwise independent channels, made possible the commercial exploitation of Blumlein's work, resulting in the 'stereosonic' system described in this paper.

This was demonstrated to members of the profession and representatives of the Press in April, 1955. In April, 1956, a full-scale public demonstration at the Royal Festival Hall was given to an audience of 1 800.

(3) COMPARISON OF BASIC SYSTEMS

In the preceding brief historical review the systems employed fall into three basic types. These are described below.

(3.1) The 'Wavefront' System

If an infinite number of microphones, placed in a vertical plane between the source of sound and the listener, were to be connected to an infinite number of loudspeakers in corresponding positions, clearly the radiated wave could be reproduced unaltered and true binaural audition would be preserved.

A line, rather than an area, of microphones and loudspeakers would give perfect location in a horizontal direction, which is adequate for most purposes since location in a vertical plane seems to be impossible except by inclining the head.

An approximation to this condition, with a corresponding limitation in the accuracy of the results, can be obtained using a finite number of microphones and loudspeakers. Contemporary film sound systems use several channels in this manner. The system described in Reference 3 also used this principle.

Fig. 1 shows the practical arrangement. Detailed tests were

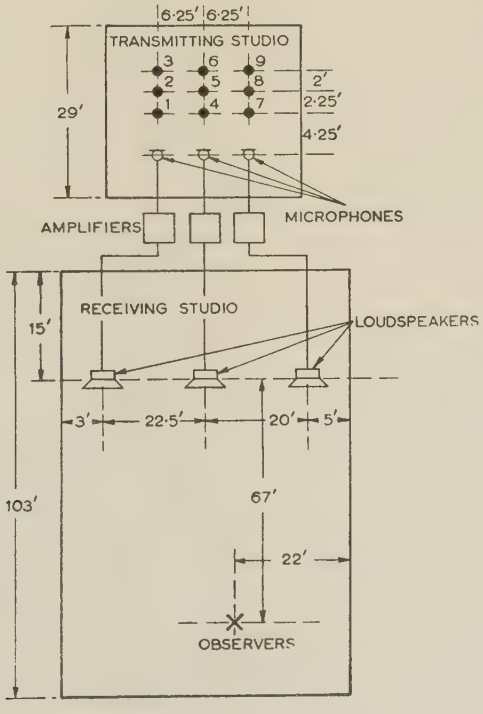


Fig. 1.—Arrangement used by Bell Laboratories in 1934.

made with a loudspeaker located at one of nine positions in the transmitting studio, and the apparent position as estimated by observers in the listening studio corresponded fairly accurately in breadth, and to a reasonable degree in depth, when the observers were not too far from the centre-line. Very-high-fidelity channels were used, and the quality of results obtained by this system was said to be better than anything ever heard prior to that date.

The use of three independent amplifier and loudspeaker channels, although suitable for public use, is quite uneconomical for domestic use. The ultimate simplification of the wavefront system to two channels can, under certain conditions, give pleasing results. For this purpose, two microphones are placed about ten feet apart in front of the sound source, driving two identical amplifying and recording or transmitting systems with two loudspeakers at a similar separation. There is a decided tendency, however, for the sounds to appear to be coming from the two separate speakers, with a gap in between in which the sound is weak.

(3.2) The Reproduction of Correct Sound Pressures at the Ears

A somewhat different approach to that of the wavefront conception is to consider local conditions at the observer's two ears. A system has been devised which relies on a consideration of these differences. The approach is based on the assumption that low-frequency sounds play little or no part in directional localization and that the principal clue to direction is the intensity difference at high frequencies produced by the shadowing effect of the head. It is shown that, if two pressure microphones are used as artificial ears in a dummy head, intensity differences between the outputs occur in accordance with the classic measurements described in Reference 3. It is then demonstrated that, if these outputs are applied to a pair of widely spaced loudspeakers placed symmetrically with respect to an observer, intensity differences do occur at his ears, although they are not as great as the original differences at the microphones. In connection with this system, attempts have been made to show that time differences are less effective than intensity differences in producing impressions of directional localization, and that incorrect time differences can be compensated by modifying the intensity differences. It is then argued that, since the major factor in localization, i.e. intensity differences at high frequencies, is reproduced, natural binaural listening is simulated.

The consideration of local conditions at the ears forms the basis of the 'stereosonic' system, described in detail in Section 4. This springs from the original Blumlein invention. It recognizes that, since each loudspeaker communicates with both ears, differences in magnitude of the sound pressures at the loudspeakers at low frequencies produce phase and not magnitude differences at the ears, since the contributions from the two loudspeakers arrive at slightly differing times. A pair of directional microphones is employed, effectively at a single position, to produce two outputs in phase, differing in amplitude according to the direction of the sound source. These are applied to a pair of spaced loudspeakers so as to produce at the ears a time difference independent of frequency at low frequencies, and an intensity difference using the shadowing effect of the head at high frequencies. It is claimed that this represents the nearest approach yet made to natural listening conditions.

(3.3) Pseudo-Stereophonic Systems

No detailed description will be given of systems in which a single microphone output supplies two or more recording channels at relative levels which are controlled manually in the dubbing stage. Such means are used in the dialogue sequences of wide-screen motion pictures and give an illusion of movement to single voices, etc., but clearly cannot give a simultaneous representation of the direction of a large number of sources as is required for most musical programmes.

(4) THEORY OF THE 'STEREOSONIC' SYSTEM

Before discussing the theory in detail, it will be advisable to review briefly what is known of the mechanism of spatial auditory location in the human being.

(4.1) Mechanism of Angular Localization

Rayleigh[4] in 1896, and Stewart[5] in 1920 both carried out experiments which demonstrated that intensity differences at the ears were insufficient to account for location at the lower frequencies, and that the phase differences had to be taken into account, although above about 1 000 c/s the intensity differences were necessary to avoid ambiguity. In spite of this and the work of others, there has been some reluctance in many quarters to accept the importance of phase differences.

Banister[6] in 1931 and the Medical Research Council[7] in 1932 give prominence to the idea that the time difference rather than phase difference may be the element detected, in which case there is no need for limitation of the effect to the lower frequencies.

In spite of this, and much other and more recent work, no exact explanation can yet be given for the mechanism of sound location. The authors believe that the two principal quantities used by an observer to estimate the angle of arrival of a sound wave are, first, the difference in time of arrival of a wavefront, and secondly, the difference in intensity at the two ears. Of these, by far the more important is the time difference. For sinusoidal waves, a constant time difference at the two ears is equivalent to a relative phase difference proportional to the frequency of the original sound. At low frequencies, where this phase difference is less than, say, π radians, the direction of arrival could be deduced from it. For such a deduction to be possible, the times of passage of each wavefront past both ears must be identifiable to the observer. As the frequency is increased, i.e. the wavefronts follow each other more closely, the point is eventually reached when a wavefront arrives at one ear before the preceding one has reached the other. Since there is nothing to distinguish between the wavefronts, ambiguities arise and it is impossible to interpret the observed time differences uniquely in terms of direction of arrival. These ambiguities start to occur when the ear spacing becomes equal to a half wavelength. At higher frequencies the head becomes an appreciable obstacle and produces an intensity difference, the magnitude of which can be used to assess the direction of arrival. It may also assist in resolving the ambiguities mentioned, and allowing the observed time differences to be interpreted at higher frequencies. Recent work in America[8] shows that the direction of arrival of pure tones can be judged unambiguously with reasonably constant accuracy up to 1 200 c/s. At this frequency the ear spacing approximates to a whole wavelength, suggesting that the first ambiguity due to confusion of wavefronts is overcome, possibly with the assistance of intensity differences. In this case the brain has to choose between two possible directions, widely separated and on opposite sides of the head.

Where the sound waveform is complex, as in the majority of natural cases, the shape of the modulation envelope probably assists in identifying the wavefronts and making possible the interpretation of observed time differences.

With pure tones at high frequencies, when it is necessary to rely on intensity differences to judge direction, accuracy deteriorates, since the amplitude differences at the ears, characteristic of the source position, may be modified by stationary waves and reflections from walls and other obstacles.

(4.1.1) Cochlear Response.

When a sound wave arrives at the ear, the immediate result is the production of an electrical waveform corresponding to that of the instantaneous sound pressure. This is known as the 'cochlear response', and it can be detected by amplifying the potential differences between suitably placed electrodes. The signal, in this form, is unsuitable for analysis by the brain, and its primary purpose seems to be to initiate an electrochemical response in adjacent nerve fibres containing the original information in pulse coded form.

(4.1.2) Action Potential.

This secondary signal is called the 'action potential' and it differs strikingly from the cochlear response. It consists of short pulses, of constant duration and amplitude, apparently occurring at random intervals. The average frequency of these pulses is related to the intensity of the original stimulus rather than its frequency. This latter is determined in some other way, possibly by observing which part of the basilar membrane is responding, although it must be admitted that pitch discrimination is more accurate than can be accounted for by simple resonance. Observations on single nerve fibres[9] show that successive pulses are always separated in time by an integral multiple of the period of the stimulus and that they occur at a particular point on the cochlear response waveform.

When a pulse has been initiated the particular fibre in which it occurs remains inactive for a short interval, known as the 'refractory period', which varies according to the strength of the stimulus. This mechanism limits the pulse rate in a single fibre to a few hundred per second. There are, however, many such nerve fibres associated with each ear, and it is known that at least one fibre will respond at each cycle of the incoming stimulus, up to a frequency of 1 500 c/s, perhaps higher.

Provided, therefore, that the brain can recognize pairs of pulses produced at the two ears by the same sound wave, the original time-difference information is available to it.

(4.1.3) Time-Comparison System.

It is clear that, if the above reasoning is correct, each cycle of the incoming wave gives rise to a pulse of action potential up to frequencies of about 1 500 c/s, and that such pulses occur at a definite point in the cycle. This happens at both ears, and if there were some means of measuring the time differences between their production, the angle of arrival of the sound could be deduced. A theory has been put forward by Jeffress[10] which suggests how this is done.

It must be noted first that transmission of a pulse along a nerve fibre is not simple electric transmission like that along a telephone cable. It is electrochemical in nature; pulses travel without attenuation as they are self-regenerating, and at a relatively slow speed. The rates of transmission as measured do not exceed about 10^4 cm/s, i.e. several times slower than the velocity of sound in air.

The theory then postulates the existence of a nerve combination such that one nerve requires to be stimulated by two others simultaneously before it will respond. Engineers will recognize this combination as being analogous to the 'logical-and gate' ubiquitous in computer circuits. It is suggested that a number of these nerve combinations are spread across the brain and stimulated by pairs of nerve fibres of appropriate length, connected to the two ears. Fig. 2 shows this schematically, in engineering terms, remembering that transmission times are proportional to the physical lengths of the different connections. A response from one of these 'gates' corresponds to a pair of pulses separated by a definite time interval arriving from the two ears. Since the lengths of the nerve fibres are significant, it is clear that the position of the nerve combination in the brain associates it with a definite time interval between stimuli, and thus with a definite direction of arrival of the original sound. In the Figure, a response from the left-hand gate corresponds to a sound wave arriving from the right side of the observer and vice versa.

Some evidence in support of this theory has recently been published. Experiments on cats are described[11] in which clicks,

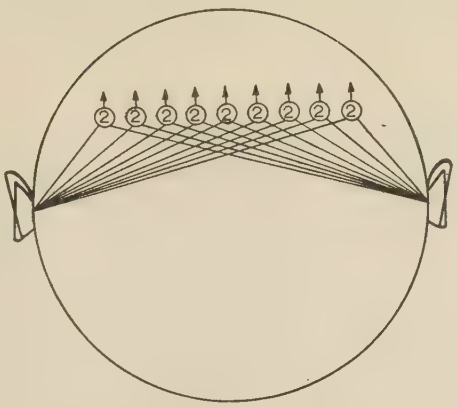


Fig. 2.—Schematic of possible time-comparison system.

separated by a known time interval, were supplied to the two ears independently and the response on the left and right lobes of the brain observed. Maximum response was obtained from the right lobe when the click supplied to the left ear was advanced by a time corresponding to a sound coming from the left of the subject, and vice versa.

(4.1.4) **Application of the Theory.**

The physiology of the brain is insufficiently known to confirm or deny the existence of such a mechanism as that postulated, nevertheless its extreme simplicity makes the idea attractive. The theory forms a useful working basis, and many of the observed facts can be fitted into it. It demonstrates clearly the importance of the first wavefront to arrive at the ears; this has the best chance of triggering a pulse, since the inhibitory mechanism is inoperative. The difficulty of localizing pure tones at high frequency is also shown. Pairs of pulses originating from different cycles of the incoming wave will produce spurious responses from the gates, which the brain will have to discard. It is thought that, under these conditions, intensity differences play an important part. With complex high-frequency waveforms, however, pulses will tend to be grouped round prominent features in the waveform, and this will again enable the time comparison mechanism to operate.

Altogether, the best chance of a clear direction seems to be at those lower frequencies where the action-potential pulses can only be paired in one way to give a time difference corresponding to a possible angle of arrival. This does not mean that the most accurate indication will be obtained at the lowest frequencies. The slower rate at which the waveform crosses the zero-pressure axis and the higher threshold as shown in the Fletcher–Munson curves combine to produce a loss of resolution that appears as an error in estimated direction. Nevertheless the sense of direction is very strong at low frequencies, and the authors do not subscribe to the view that low frequencies play no part in angular localization.

(4.2) The Aim of the 'Stereosonic' System

The aim of the 'stereosonic' system is to reproduce at the ears of the listener, in as large an area as possible in front of a pair of loudspeakers, the same vector sound pressures as he would have experienced by direct listening in a corresponding position in front of the sound stage. In other words, although the spacing between the loudspeakers must, in general, be less than the width of the original sound source (except for some solo instruments and unaccompanied solo voices), if the apparent angular width of the reproduced sound is the same as that subtended at the optimum position for direct listening, the listener will imagine that he has been conveyed to the optimum position for direct listening in the recording studio. So far as is known, the ratio of reverberant to direct sound provides the only clue to distance of the source, except for any possible change in quality due to absorption of high frequencies at a distance. In general, the reverberation of the average domestic room is small compared with that of concert halls used for large musical combinations, and so the sense of distance thus conveyed is not substantially altered.

It is impossible to achieve this aim with perfection, but an attempt is made in the system to reproduce the actual relative phases and thus the inter-aural time differences at the lower frequencies and the relative intensities at the upper frequencies.

It should also be clearly understood that the system is intended to operate only with spaced loudspeakers and thus transmit the spatial effect when listening in free space. The system will not function correctly if the two channels are connected to left and right headphones.

(4.3) Mathematical Theory

(4.3.1) **Vector Pressures at Ears in Direct Listening.**

Fig. 3 shows an actual source of sound S before two ears E_L and E_R, spaced by a distance h in such a position that its

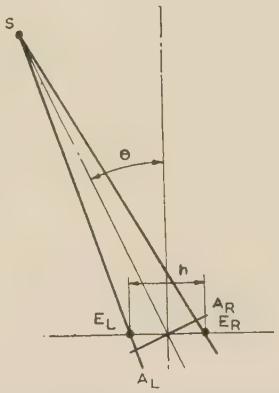


Fig. 3.—Time differential of sound at two ears from oblique source.

direction is at an angle θ to the face-on position. The sound to E_R, however, will travel a distance E_RA_R further than, and that to E_L a distance E_LA_L less than, the average. If v is the velocity of sound in air, the sound will take a time $(h \sin \theta)/2v$ to travel from E_L to A_L. Thus the time interval between the arrivals of sound at the two ears will be $(h \sin \theta)/v$. Hence, if h is small compared with the distance from the source, the magnitude at each ear will be the same, but there will be a phase difference

$$\phi = \frac{\omega h \sin \theta}{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where ω is 2π times the frequency of the sound wave.

At higher frequencies, when the wavelength is short compared with h, the phase angle will be large and ambiguous but the time delay will be the same. Owing to the masking effect of the head, the magnitude is not subject to accurate calculation but has been determined experimentally.[12, 13]

If these phase differences and magnitudes can be simulated in reproduction, the received sound will appear to come from the same angle θ.

(4.3.2) **Reproducing System.**

For domestic purposes it is generally permissible to use two loudspeakers only, and they must be operated in such a way as

to reproduce as nearly as possible the required vector pressures at the ears. Since the bulk of acoustic energy in music and speech occurs below about 700 c/s, it is important that the system should operate effectively at low frequencies.

In Fig. 4, W_L and W_R represent two similar loudspeakers

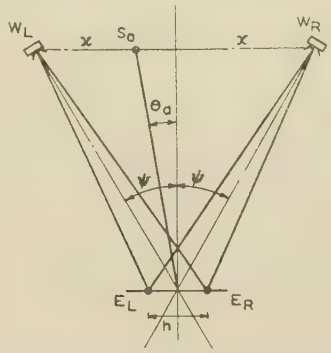


Fig. 4.—Apparent source deduced from the time differential of sound produced at two ears by two loudspeakers operating at different sound levels.

supplied with inputs which are in phase but of different magnitudes. The loudspeakers subtend an angle 2ψ at a centrally placed listener whose ears E_L and E_R are separated by a distance h, which is assumed to be small compared with the loudspeaker distance. From eqn. (1) the phase difference between E_L and E_R for sound coming from either speaker will be $2\omega\mu = (\omega h \sin\psi)/v$, where 2μ is the difference in time of arrival at E_L and E_R.

Let the average instantaneous sound pressure at both ears be $L \sin \omega t$ from W_L and $R \sin \omega t$ from W_R. The sound pressures at each ear can be calculated as follows:

	Average	*At* E_L	*At* E_R
From W_L	$L \sin \omega t$	$L \sin \omega(t+\mu)$	$L \sin \omega(t-\mu)$
From W_R	$R \sin \omega t$	$R \sin \omega(t-\mu)$	$R \sin \omega(t+\mu)$

Total pressure at $E_L = L \sin \omega(t+\mu) + R \sin \omega(t-\mu)$

$$=\sqrt{(L^2+R^2+2LR\cos 2\omega\mu)}\sin\left[\omega t + \text{arc tan}\left(\frac{L-R}{L+R}\tan\omega\mu\right)\right]$$

Similarly, the total pressure at E_R is

$$\sqrt{(L^2+R^2+2LR\cos 2\omega\mu)}\sin\left[\omega t - \text{arc tan}\left(\frac{L-R}{L+R}\tan\omega\mu\right)\right]$$

Hence the phase difference between E_L and E_R is ϕ_2,

where $$\phi_2 = 2 \text{ arc tan}\left(\frac{L-R}{L+R}\tan\omega\mu\right)$$

$$= 2 \text{ arc tan}\left(\frac{L-R}{L+R}\tan\frac{\omega h\sin\psi}{2v}\right)$$

When $\omega\mu$ and $\phi_2 \ll \pi/2$

$$\phi_2 = \frac{L-R}{L+R}\frac{\omega h\sin\psi}{v} \quad . \quad . \quad . \quad . \quad (2)$$

Thus, if the loudspeakers are supplied in phase with correct relative amplitudes, the phase difference ϕ_2 at the ears can be made to be the same as that from a sound source at any angle within $\pm\psi$. In addition, if L or R is permitted to become negative, i.e. to have a phase reversal, the apparent direction lies outside the limit of $\pm\psi$ at frequencies where this analysis is valid. This phenomenon is not readily observable in practice possibly because it is accompanied by a certain degree of amplitude cancellation. It has, however, been observed under laboratory conditions.[8]

(4.3.3) **Microphone Systems.**

When Blumlein's first experiments were carried out in 1929–30 the only readily available microphones were pressure-operated and had substantially circular polar diagrams. In order to obtain the required inputs to the loudspeakers the first arrangement to be used consisted of two such pressure microphones separated by about 8 in (i.e. typical distance between ears). The outputs were therefore the same in magnitude, for any source angle, but there was a phase difference given by eqn. (1). To convert this into the required amplitude difference demanded by eqn. (2), an ingenious circuit was used in which the two microphone outputs were first summed and differenced to produce two new voltages in quadrature. The difference voltage was integrated, i.e. multiplied by a factor proportional to $1/\omega$ and rotated by 90°. The sum voltage was uniformly attenuated by a suitable amount. These two modified voltages were again summed and differenced, giving two final voltages, in phase but of different amplitudes, which it can be shown are of such values that, when applied to two loudspeakers, will, according to eqn. (2), produce at the ears the same phase angles as those at the microphones.

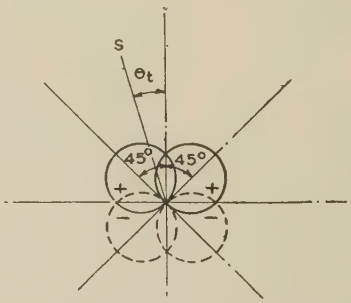


Fig. 5.—Polar characteristics of a pair of velocity microphones at 90°.

Fig. 5 represents two velocity microphones, e.g. ribbon type, with their axes of maximum response at 90°. Such microphones have a response which is proportional to the sine of the angle between the source and the plane of the ribbon. A source S_t is assumed to be at an angle θ_t from the median axis of the microphones. The outputs from the microphones will be in phase at all values of θ_t if the microphones are placed on the same vertical axis but their outputs L and R are given by

$$L \propto \sin(45° + \theta_t)$$
$$R \propto \sin(45° - \theta_t)$$

whence it can be shown that

$$\frac{L-R}{L+R} = \frac{\sin\theta_t}{\cos\theta_t} = \tan\theta_t \quad . \quad . \quad . \quad . \quad (3)$$

(4.3.4) **Performance of Complete System at Low Frequencies.**

If the outputs of the microphones are connected (after suitable amplification) to the loudspeakers, the resultant phase difference between the sounds at the ears will be

$$\phi_2 = \frac{L-R}{L+R}\frac{\omega h\sin\psi}{v} \quad . \quad . \quad . \quad . \quad (2)$$

$$= \tan\theta_t \frac{\omega h\sin\psi}{v} \quad \text{from eqn. (3)}$$

The ears will interpret this as an apparent sound source at an angle θ_a, where $\sin\theta_a = \phi_2 v/\omega h$, from eqn. (1);

i.e. $$\sin\theta_a = \tan\theta_t \sin\psi \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Fig. 6 shows the apparent angle plotted against the true angle for various values of ψ. It will be seen that, when the listener

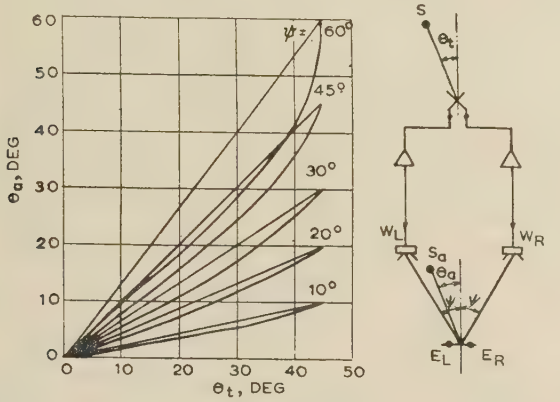


Fig. 6.—Apparent angle plotted against true angle for velocity microphones at 90°.

is at such a distance that the loudspeakers subtend an angle of 120°, the apparent angle is very close to the true angle up to ±35°. When the listener is at other distances from the loudspeakers, but still on the centre line, the apparent source remains at the same fraction of the total loudspeaker spacing, thus showing that a correctly proportioned sound picture is presented although the angular scale has been altered. The angular distortion is such that the apparent source appears to be somewhat nearer the centre than it should be over most of the range.

(4.3.5) Performance at High Frequencies.

If the velocity microphones have a constant polar characteristic at all frequencies, the ratio of voltages applied to the loudspeakers is independent of frequency. At frequencies above about 700 c/s, however, the foregoing analysis will not hold, since the phase angle between the pressures at E_L and E_R will be ambiguous. At these frequencies the masking effect of the head, however, causes the left ear to be more affected by the left loudspeaker than by that on the right and vice versa, and it can be easily demonstrated experimentally that directional information can be conveyed at the higher frequencies.

Subjective tests were made with a number of observers using two loudspeakers supplied with known relative voltages from a source of recorded music. First a filter was inserted passing all frequencies up to 700 c/s from a variety of sources of sound including male and female speech, solo, orchestral and brass-band music. The experiment was repeated using all frequencies above 600 c/s. Quite definite location within about ±2° was obtained in each case, but whereas at low frequencies the angle was in agreement with that predicted from eqns. (1) and (2) for a given loudspeaker ratio, that obtained at high frequencies was greater. The relationship obtained is in agreement with that published by other workers who rely primarily on intensity differences.[15] By introducing a factor of approximately 0·7 into the ratio $(L - R)/(L + R)$ above 700 c/s, the results for high and low frequencies can be brought into line, except for extreme positions of the source.

(4.4) Basis of Recording System

To meet the requirements set out in Section 4.2, therefore, the microphone system can consist of a pair of velocity microphones with mutually perpendicular axes of maximum response. The outputs from these microphones can be summed and differenced and the difference channel subjected to a loss of 3 dB at all frequencies[14] above 700 c/s. These voltages are then again summed and differenced and used to supply two identical loudspeaker channels. If the loudspeakers are equidistant from the observer and subtend an angle of approximately 90° at him, he will receive substantially correct directional information at all frequencies within the spectrum. At greater distances from the loudspeakers the listener will observe the same apparent position of the source, relative to the loudspeakers.

An important feature of a system using two coincident velocity microphones at 90° is that the r.m.s. sum of the outputs of the loudspeakers is constant for a source at a constant distance from the microphones, regardless of its direction. It is this feature which enables the system to reproduce a uniform sound field between the loudspeakers without the tendency to leave a 'hole' in the centre. This latter effect can be very pronounced with the spaced microphone system.

(4.5) Sources of Error

Although the two amplifying channels and the microphones and loudspeakers have been assumed to be distortionless in the above explanation, it is clear that they cannot be perfectly matched in amplitude and phase, particularly the transducers. The effect of the small inevitable differences in frequency response can be calculated from the preceding theory, but such transducers will, in general, introduce phase shifts before amplitude differences become very pronounced.

Calculations show that a phase shift in one channel produces no change in the central position, i.e. when the two channels are at the same level. A phase shift in one channel of as much as 90° can introduce error factors of 2 at small values of θ_t, 1·5 at $\theta_t = 26°$ and 1·2 at $\theta_t = 39°$.

(4.5.1) Effect of Asymmetric Listening Position.

The preceding calculations have assumed an observer symmetrically placed with regard to the loudspeakers and facing the centre line. If the observer turns his head the apparent source will appear to remain in the same position, as in natural listening, but if he moves to either side of the centre line the apparent position will move in the same direction, e.g. from S_a to S'_a in Fig. 7.

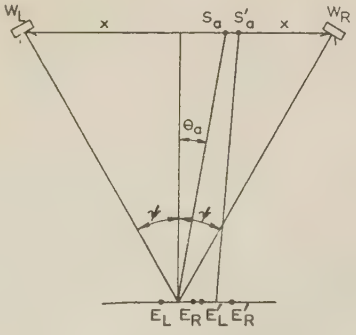


Fig. 7.—Change of position of apparent source with displaced listening position.

The reasons for this are, first, that the loudness from W_R has increased and that from W_L decreased, and their relative phases are altered owing to the difference in the path lengths. Secondly, the angles subtended by the loudspeakers at the observer have also become unequal. The effect of these

inequalities makes the calculation of apparent angle exceedingly laborious, but it is possible, by making simplifying assumptions, to appreciate their general effect. Fig. 8 shows the position of the apparent source of Fig. 7 for various listening positions plotted against the positions of the real source with respect to the microphones. In Fig. 8, r is the projection of the observer's

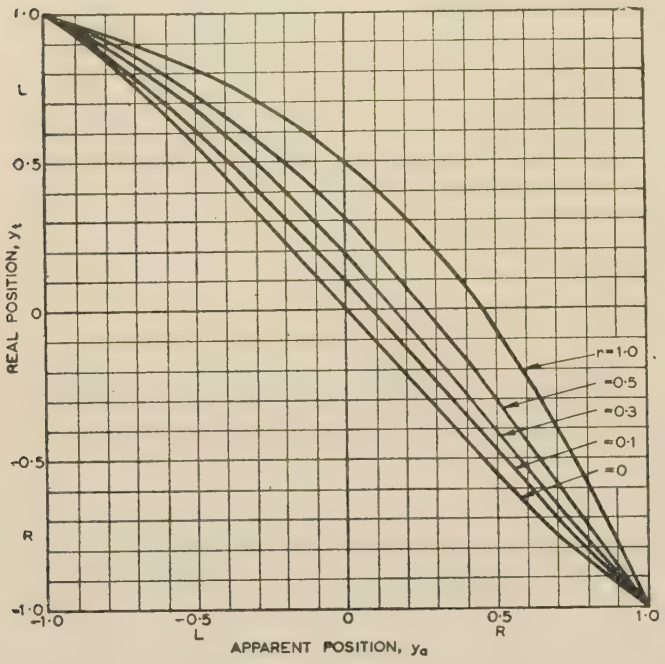


Fig. 8.—Apparent position plotted against true position for asymmetric listening position.

displacement on the loudspeaker base-line, expressed as a fraction of x. The calculations ignore the effect of the path difference on the relative phases at the listener of the sound from W_L and W_R; i.e. it applies where such a path difference is small compared with a wavelength. The curves bear out the statement that the apparent source moves in the direction of the observer.

In order to take account of the effect of path differences it is necessary to fix the scale of the layout shown in Fig. 7, so that a definite phase shift can be associated with any given listening position at a particular frequency. For this purpose let d equal 10 ft to represent a typical domestic arrangement. A listening position is chosen corresponding to $r = 0{\cdot}5$, and the apparent position is plotted against frequency in Fig. 9 for various values of the true angle θ_t. It will be seen that, as the frequency increases from zero, deviations from Fig. 8 are initially small, but they increase rapidly above 100 c/s, reaching a maximum in the region of 200 c/s, beyond which the function approaches its initial value once more before reaching a second maximum at about 600 c/s and again returning to its original value.

The calculations have not been made for frequencies above which the theory given in Section 4.3.2 is valid.

The function may reach a finite maximum, which in all cases is seen to lie outside the loudspeaker base-line, or it may become imaginary. In this case, the phase difference at the ears is so large that it cannot be interpreted in terms of any real angle of arrival. These regions are the ones in which the path difference approximates to an odd number of half wavelengths, the sound from the two loudspeakers arriving substantially out of phase at the two ears.

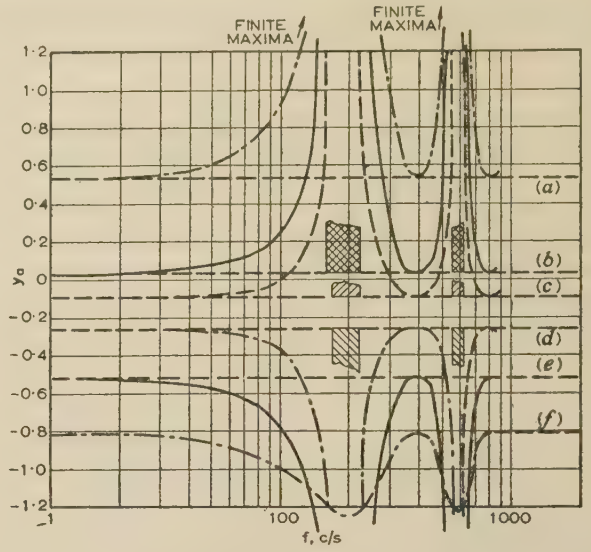


Fig. 9.—Apparent position as a function of frequency for asymmetric listening position.

(a) $\theta_t = 36{\cdot}9°$.
(b) $\theta_t = 18{\cdot}4°$.
(c) $\theta_t = 11{\cdot}3°$.
(d) $\theta_t = 0$.
(e) $\theta_t = -18{\cdot}4°$.
(f) $\theta_t = -36{\cdot}9°$.

Fortunately, at these critical frequencies the resultant magnitude becomes small, thus allowing the remaining sound, which has small phase errors, to predominate. The importance of scale is obvious from the curves; halving all dimensions will allow the system to operate to twice the frequency before the anomalous regions are approached.

The effect of the offset listening position is worst for true angles near the centre, when L and R have about the same amplitude. (The extreme positions, $\theta_t = \pm 45°$, are unaffected, since only one loudspeaker is operating.)

(5) RECORDING EQUIPMENT

(5.1) Microphone Combinations

(5.1.1) Pressure Microphones.

As stated in Section 4.3.3, the earliest experiments were necessarily carried out with pressure microphones. These had substantially spherical polar diagrams—at least, at low frequencies. The problem of producing magnitude differences was solved by giving them a small but definite spacing, summing and differencing their outputs, and integrating the difference vector before recombining to drive the loudspeakers. The significant parameter in such a case is the actual spacing employed. If this exceeds about a quarter wavelength of the highest working frequency, the resultant outputs to the loudspeakers become of opposite polarity and the system breaks down. Small spacing is desirable on this score. However, if the spacing is too small, the difference vector will diminish until it is comparable with the normal random differences in the microphones. Moreover, the amplification in the difference channel has to be increased to compensate, in addition to having a characteristic inversely proportional to frequency as already mentioned. Low-frequency electrical noise introduced by the amplifiers thus becomes the factor limiting the reduction in microphone spacing. A way out of this difficulty was sought by using more than one pair of microphones to cover the required frequency band, but, as can be imagined, this arrangement was cumbersome in practice and complicated theoretically. Pressure microphones were finally

abandoned after a composite microphone system had been tried, using two pairs of crystal elements, one pair being spaced 8 cm apart and operating up to 1 000 c/s and the other being spaced 1 cm apart and operating from 1 000 c/s upwards.

(5.1.2) **Velocity Microphones.**

By this time, ribbon microphones were becoming available. As is well known, these have a cosine-law polar characteristic in the horizontal plane. A pair of these can be used in two ways to produce the required outputs for driving a pair of loudspeakers. For example, they may be used with one microphone having its axis directed to the centre of the sound stage and the other as close as possible but with its axis at right angles. In this condition the outputs from the microphones will be in phase and invariant with frequency (in so far as their characteristics are flat), and will correspond to the sum and difference vectors described in Section 5.1.1 after manipulation but before recombination to drive the loudspeakers.

Alternatively, they may have their axes equally inclined to the centre of the sound stage. In this condition their outputs are suitable for direct amplification and reproduction by a pair of spaced loudspeakers. Such outputs may be considered to be of 'left and right' type rather than 'sum and difference', as in the previous instance.

In either case it will be noted that the operation of integration that had to be performed in the case of pressure microphones is unnecessary. Consequently, although ribbon microphones are generally less sensitive than pressure types, the overall signal/noise ratio of the system is improved.

(5.1.3) **Other Types of Microphones.**

It will be apparent that any pair of microphones having a polar characteristic other than circular can be used in this way to produce 'left and right' type signals of different amplitudes at a pair of loudspeakers. Whether these combinations are of value depends on the uniformity with which their polar characteristics can be maintained with frequency, as well as the degree within which the microphones can be matched in amplitude and phase.

Combinations of microphones having dissimilar polar characteristics may also be employed to produce 'sum and difference' type outputs, provided that their amplitude and phase characteristics are well matched. Combinations of polar diagrams which appear to show promise are those of circular with cosine, which would give the same directional response as the pressure microphones used in the early experiments, and cardioid with cosine, which would give substantially a one-sided version of the same thing.

(5.1.4) **Current Practice.**

The combination of two cosine microphones has so much to recommend it compared with other known arrangements that the principal effort has been concentrated on its improvement. If a pair of ribbon microphones is used, or even a specially-designed double-element ribbon type, the difficulties encountered are still quite serious. At low frequencies the fundamental resonances of the two ribbons cause relative phase shifts unless these are adjusted to occur at the same frequency and to have the same degree of damping, whilst at high frequencies the ribbon microphone having the theoretical cosine-law polar diagram has yet to be designed. Departure from constancy of polar diagram with frequency will upset the requirement that the amplitude differences for any given angle of arrival shall be independent of frequency. Nevertheless, with all their limitations, some very acceptable recordings have been made using ribbon types.

Within recent years the condenser microphone has undergone considerable development, enabling directional characteristics to be obtained from a double-diaphragm element of small size. Since these elements are substantially free from mechanical resonances within the audible spectrum, their amplitude and phase characteristics show great uniformity. This is fortunate as the polar characteristics show some variation from sample to sample, particularly at low frequencies. Selection has thus to be principally on the basis of polar response. Individual adjustments to match their sensitivities can easily be made. Pairs of these microphones, mounted with their elements as close together as possible in a single cylindrical holder, have been used with some success. Like all arrangements with one element above the other it is necessary to ensure that the common axis is perpendicular to the plane in which the various sounds lie, otherwise the vector relationships are upset.

(5.2) Studio Techniques

The problems encountered in making a 'stereosonic' record are not all of an engineering nature. The finished record must have a pleasing tonal balance, and the apparent spatial distribution of the instruments, if not of prime importance, must not sound unnatural. Using a single pair of microphones, it is often extremely difficult to satisfy the required conditions. In single-channel recording, good results have sometimes been achieved using only one microphone, but the occasions when this is possible are the exception rather than the rule. Generally, in order to achieve a proper balance, multi-microphone techniques have to be adopted. For obvious reasons this cannot be done when the direction of the various sources is significant. It is therefore necessary to rely on correct disposition of the performers in the recording studio to achieve the required tonal balance. In this connection, it may be remarked that the double-cosine combination has two working arcs each 90° wide on opposite sides of the common axis. This has been found useful in practice, particularly for large choral works, in which the orchestra has been arranged on the one side, with choir and soloists on the other. The adoption of such unusual layouts, however, often makes the task of the conductor more difficult, since he may not be able to see all the performers without turning round. Considerations of this sort, which may be termed the technicalities of musical performance, often preclude the use of the purely engineering solution and make the task of the recording engineer even more complicated.

Under difficult circumstances it is possible to envisage the need for more than one microphone pair, but it is realized that the use of additional microphones may introduce as many problems as it solves.

(5.3) Microphone Amplifiers and Equalizers

Amplifier design follows conventional lines. It is, of course, essential that pairs of amplifiers used in this type of recording have accurately matched frequency and phase characteristics, and that their gain be constant over long periods. In general, however, modern designs employing negative feedback have no difficulty in achieving the required consistency. The same requirement for accurate matching applies to any microphone equalizers that are used, e.g. for bass correction when working in close proximity to an artist.

(5.4) Mixers

As stated in Section 5.2, the use of a single microphone pair for stereophonic recording is almost mandatory. It is nevertheless possible to foresee circumstances in which it may be necessary to use a second pair or to introduce extraneous effects.

Some form of mixer is therefore necessary, although this will generally be a simple device compared with those used in conventional recording, which may have to cater for relatively large numbers of microphone channels.

A stereophonic mixer in current use does permit the use of two microphone pairs with independent level controls. It also includes means for providing the correction at high frequencies referred to in Section 4.4. This requires that the signals, if not of the sum-and-difference type, shall have been subjected to a sum-and-difference operation.

A further facility is provided which gives an additional degree of flexibility in the recording studios. This consists of a differential control permitting the relative amplification of the 'sum and difference' channels to be altered over a limited range. This control allows the apparent angle of the total reproduced sound to be modified and can be used to good effect in several ways. The microphones may have to be withdrawn from the artists in order to get more 'ambience' into the recording. This will reduce the apparent stage width, but the full width can be restored by suitably attenuating the sum channel relative to the difference. Similarly, if the need for great 'presence' calls for a very close microphone position, the reproduction may cause a solo instrument to sound much too large, and this can be corrected by attenuating the difference channel relative to the sum.

The signals are finally subjected to a second sum-and-difference operation to convert them to 'left and right' type for application to the input of a twin-track tape recorder. The replay amplifiers of the machine drive two matched loudspeakers suitably adapted for stereophonic monitoring.

A simplified diagram showing the operations performed in a typical recording channel is given in Fig. 10.

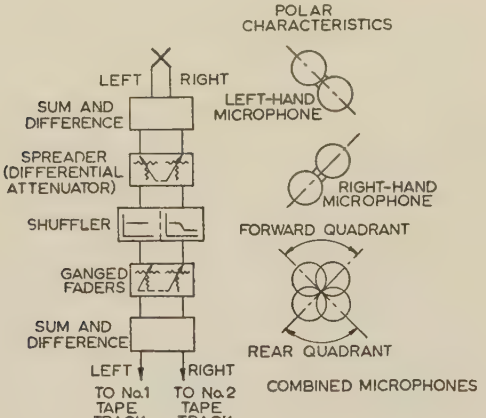


Fig. 10.—Functional schematic of recording system.

An interesting feature of the control system is the peak-level indicator. The use of two such meters, one on each channel, is unsatisfactory owing to the difficulty of observing them simultaneously. In this equipment a peak-level indicator circuit,[16] with the ability to measure the true level of short transients, is connected to each recording channel, and the meter on the mixer panel can be switched to either channel at will. Alternatively, for general use, a circuit is provided whereby the meter is electronically switched to whichever channel has the higher level. The operator then uses his main gain control in the ordinary way to avoid overloading either channel on the tape.

(5.5) Magnetic-Tape Recording Machines

In normal record production it is the practice, irrespective of the final form of the commercial article, to make master recordings on magnetic tape for convenience of play-back, editing, etc. These advantages apply equally to stereophonic recording. The system described employs professional machines of high quality, modified by the addition of twin-track heads and twin-head amplifier channels.

A standard tape speed of 15 in/sec has been chosen for two reasons:

(*a*) Satisfactory standard of master quality can be obtained at this speed.

(*b*) The majority of single musical works can be recorded within the full 11 in-diameter tape spools and thus simplify the copying process.

The record heads are provided with bias from a common oscillator to avoid heterodyne beats due to residual crosstalk.

(6) RECORDING AND REPRODUCING METHODS

(6.1) Choice of Media

It is essential that the recording of the two channels should be made on a medium which allows the two tracks to be permanently linked together so that the correct phase relation between the signals is maintained throughout any copying process and replay operation.

In 1931 when Blumlein conceived the basic idea of his system, the only recording medium capable of giving reasonable results was the wax-cut disc with its electrolytic copying process. He therefore proposed applying the two signals from the microphone system to the complex transducer so arranged as to cut a single groove. If the axes of movement of the transducer armatures carrying left and right signals are inclined at 45° to the surface of the wax and are thus at 90° to each other, the resulting lateral cut can be arranged to represent the sum of the two channels and the 'hill and dale' to represent the difference. Such a disc could be played on a normal gramophone to give the equivalent reproduction of a single-channel recording. The early experimental samples of the complex-cut disc suffered from considerable interaction at the higher frequencies and from excessive background noise. The fine-groove technique may simplify the problem, but to produce a complex-cut disc to the standards demanded by the modern record is very difficult, and the problem has not yet been solved satisfactorily.

In the United States stereophonic disc records have been marketed in which the lateral-cut tracks are disposed in two concentric bands. This record is replayed with two pick-up heads mounted on a common arm. The operation of locating both pick-up styli in their correct grooves is a delicate one. The frequency response and distortion characteristic of the two tracks will differ quite appreciably when there is a large difference between the track diameters.

There have been various suggestions for modifying the frequency band of the signals of one of the channels, so that both channels can be recorded on one groove by means of a single transducer. One such method proposed by Livy[17] suggests that one set of signals is arranged to modulate a carrier, then to select the lower sideband and apply this, together with the normal signal band of the other channel, to a wide-range cutter head. Reproduction is effected by using a wide-range pick-up, separating the two frequency bands and applying the upper band to modulate a carrier of the same frequency as that used for recording. The lower sideband produced by this modulator contains the same frequency components as the original signal. Livy also proposed to record the carrier frequency on the disc so that it would be used during reproducing to control the frequency of the local oscillator. This double-bandwidth system demands a very high performance in terms of frequency response, and thus if a frequency range up to 10 kc/s be required, the full bandwidth

of the system will be at least 20 kc/s. This wide bandwidth is very difficult to attain, particularly at the inner groove diameters.

If recordings are made side by side on a continuous film, the synchronization problems are largely overcome, and while it is recognized that optical methods have been used, the recent development of the magnetic-tape process makes it a more attractive system.

(6.1.1) **Crosstalk.**

The overall crosstalk between channels should be better than 30 dB at all frequencies. In the production of a commercial tape record there are at least four stages where crosstalk can take place:

Master recording.
Master replay.
Copy recording, commercial tape record.
Copy replay.

If copy masters are used two more stages of crosstalk are introduced. In practice, it has been found that the same amount of crosstalk is introduced during the recording process as during the reproducing process.

In well-designed heads a crosstalk of −50 dB can be attained, and if the crosstalk at the various copying stages can be added arithmetically the overall figure for a commercial tape record will be −38 dB, or if a copy master is used, −34 dB. If it is required to replay either twin-track or single-channel half-track tape records from the same head system, the crosstalk of this replay head must be better than −55 dB, and preferably better than −60 dB, in the 1 000–3 000 c/s region.

Crosstalk can take place owing to the mutual inductance of the heads and owing to leakage fields from one track to the head on the other track. The mutual inductance can be reduced by placing magnetic screens between the heads and separated from them by brass or aluminium spacers. These screens should be large enough to shield the magnetic circuit and the windings. The usual back gap should be eliminated, and the head dimension should be kept as small as possible.

A useful reduction of crosstalk can be achieved by either series or parallel cross-connection of the windings, so that a small signal from one head is injected in anti-phase to the leakage signal on the other head. If carefully adjusted the series cross-connection can reduce the residual crosstalk by at least 10 dB.

Appreciable crosstalk may occur if the front edge of the outer magnetic screening shield is too close to the magnetic tape. The flux from one track enters the shield and then passes to the magnetic yoke of the other head operating on the other track. A clearance between the tape and the shield of 0·2 in is sufficient to avoid this effect.

(6.2) Track Standards

The early experimental stereophonic magnetic-tape recordings were made with the heads displaced by $2\frac{1}{16}$ in. This enabled the conventional half-track heads to be used on the existing recording and replay machines, and the separation between the heads was sufficient to make crosstalk negligible. When copy tapes were made for issue as commercial records it was soon obvious that the use of in-line heads would make for easier acceptance as an international standard for domestic and commercial tape records. It was also decided to change to in-line heads on the master recordings, so that the process of editing was simplified. Furthermore, the use of in-line heads eliminates phase shift between the signals on the two tracks owing to small variations of elasticity of the tape along its length.

Standardization has been effected on the designation of the tracks, which is as follows:

If the tape moves from left to right and with the active side facing away from the observer the top track shall be designated *No.* 1 *track* and shall carry the recording for the left-hand channel as viewed from the audience. The bottom track shall be designated *No.* 2 *track* and shall carry the recording of the right-hand channel.

The replay response characteristic (100 microsec) and the track dimensions are in accordance with Amendment No. 1 to B.S. 1568: 1953, relating to tape speeds of $7\frac{1}{2}$ in/sec. The track dimensions are as shown in Fig. 11.

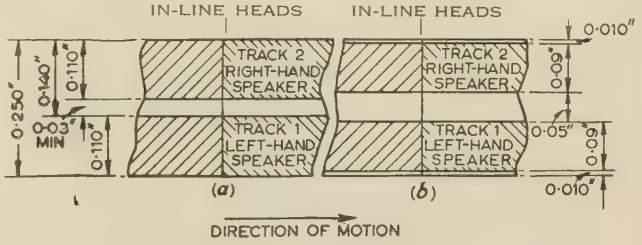


Fig. 11.—Track standards.
(*a*) Record-head track.
(*b*) Replay-head track.
Views looking on coated side of tape.

(6.3) General Requirements for Domestic Operation

The tape transport mechanism may be of the conventional type, but the associated twin-track magnetic replay head must conform to the standards laid down above. Since the signals from the two tracks on the tape are left and right, it is essential that the gains of the two replay channels should be closely matched, and for this purpose preset controls should be provided for initial adjustment. It is a further requirement that the equality of gain should be maintained for all settings of the main coupled gain control. Large relative phase shifts in the two channels should be avoided, particularly at low frequencies.

The underlying principle of the system demands that the loudspeakers should possess, as far as possible, uniform polar response characteristics in the horizontal plane at all frequencies. It has been suggested that departure from uniformity can be an advantage in maintaining the apparent position of the source for a wide range of listening positions. Such non-uniform polar characteristics cannot reasonably be attained except at high frequencies. Even if it were possible to extend the directional characteristics down to the low frequencies, the beneficial results claimed would not be realized on account of the path differences, as outlined in Section 4.5.1. At the high frequencies non-uniform polar characteristics have the disadvantage that the overall tonal balance will vary with the listener's position. Furthermore, the directional response will emphasize any background hiss and will tend to identify the two loudspeakers as separate sound sources and thus interfere with the illusion. In addition, the domestic user might have difficulty in positioning the loudspeakers with sufficient accuracy.

If cone loudspeakers are used the rear radiation should be suppressed, and this can be conveniently done by using a closed box baffle. Such baffles can have quite small volume, and they can be designed to occupy a very small floor area. Any loss of bass response due to the small enclosed volume can be compensated electrically in the power amplifiers.

(6.3.1) **A Commercial Model.**

One form of domestic reproducer has been described by Smith and Martin,[18] but for the sake of completeness a brief general description may be of interest. The complete machine consists of two consoles each containing a loudspeaker group and a power amplifier. One of these consoles also contains a tape deck with its associated head amplifiers, tone and gain controls. Fig. 12 is a schematic. Bass and treble ganged tone

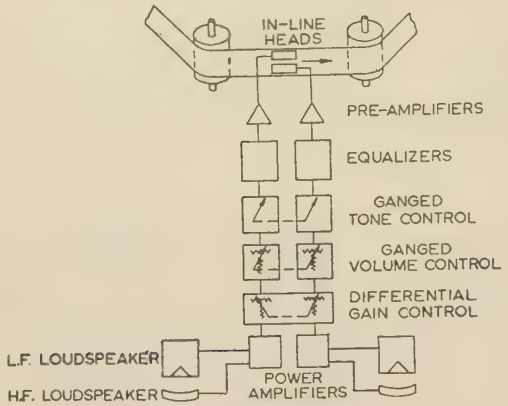


Fig. 12.—Functional schematic of reproducer.

controls are provided, in order to allow some adjustment for the different acoustic conditions which may be met.

In order to maintain uniformity between the two channels the controls are of the stepped-switch type. The main gain control consists of two ganged switches with 12 steps of 3 dB each. There is a differential gain control on the panel to take up any slight drift in amplifier gains and to allow for asymmetry in the listening room. A preset continuous control on the amplifier chassis balances the gain during factory adjustment. The power amplifiers are rated at a peak output level of 10 watts.

In order to give the best horizontal polar distribution the elliptical cone loudspeakers are mounted in closed rigid box baffles of $3\frac{1}{2}$ ft^3 capacity, the major axis of the cone ellipse being vertical. The frequency range of this speaker is limited at the upper end to about 5 000 c/s, in which region the electrostatic speaker takes over and continues beyond 15 kc/s. The electrostatic speaker consists of a curved metal back plate 24 in long by $1\frac{1}{4}$ in wide, over which is laid a membrane metallized on the side away from the back plate. This speaker is driven from a separate amplifier and is supplied with a bias potential of 300 volts.

(6.3.2) Domestic Listening Conditions.

In general, 'stereosonic' tape records are balanced with regard to tonal quality and perspective, for reproduction in the average home lounge in which the reverberation time is of the order of 0·5 sec or less. This reverberation period is short compared with that in the studio or concert hall in which the recordings will have been made; it is not likely to interfere with the listener's impression of the original studio conveyed to him by the stereophonic effect; and there will be no confusion due to the two distinct reverberations which might blur the detail. Provided that the walls and the floor are reasonably absorbent it is found that, in spite of errors due to asymmetry, one can move about over a large floor area in front of the loudspeakers without losing the major benefits of this type of reproduction. As stated earlier in the paper, the optimum position for listening is at the apex of an equilateral triangle with the base-line formed by the line joining the two loudspeakers; this is the arrangement used for monitoring during recording. In the average room it is satisfactory to set the loudspeakers 10 or 12 ft apart, but in a restricted space a very fair performance can be obtained down to distances of as little as 4 or 5 ft.

(6.4) Operation of the System in Large Halls

Special difficulties attend the reproduction of 'stereosonic' records on a large scale, apart from the general fact that, like all other types of record, they are balanced primarily for domestic conditions.

The most serious errors are caused by increased path differences. These occur, as outlined in Section 4.5, down to lower frequencies and over a much greater proportion of the listening area than is the case when operating at the scale for which the system was designed.

In a small room reflections from the walls, etc., decay rapidly and follow each other at such short intervals that the resulting sounds appear to the listener to coalesce. In large halls the relatively undamped primary reflections from the surroundings arrive at the listeners' ears at intervals sufficiently great to cause an appreciable distortion of the sound picture.

Ideally the system requires that the loudspeakers shall be uniformly radiating point sources. In the domestic case a reasonable approximation to this can be achieved because single small units can be constructed to handle the required power. In large-scale reproduction, multiple units or single units of appreciable size are necessary; this departure from the ideal can cause a depreciation of the stereophonic definition.

(7) ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance of many members of the technical staff at the Recording Studios and the Research Laboratories of Electric and Musical Industries, Ltd., and Messrs. D. E. Broadbent and B. Shackel for helpful discussion on the physiology of hearing.

They also wish to thank the Directors of Electric and Musical Industries, Ltd., for permission to publish the paper.

(8) REFERENCES

(1) 'The Telephone at the Paris Opera', *Scientific American*, 31st December, 1881.
(2) Blumlein, A. D.: British Patent No. 394325.
(3) Fletcher, H., *et al.*: 'Auditory Perspective', *Bell System Technical Journal*, 1934, **13**, p. 239.
(4) Lord Rayleigh: 'Theory of Sound, Vol. II', p. 440.
(5) Stewart, G. W.: 'Binaural Location of Pure Tones—I and II', *Physical Review*, 1920, **15**, p. 425.
(6) Banister, H.: 'The Basis of Sound Localization' (Physical Society Discussion on Audition, June, 1931).
(7) 'Studies in the Localization of Sound' (Medical Research Council Report, 1932).
(8) Sandel, T. T., Teas, D. C., Feddersen, W. F., and Jeffress, L. A.: 'Localization of Sound from Single and Paired Sources', *Journal of the Acoustical Society of America*, 1955, **27**, p. 842.
(9) Galambos, R., and Davis, H.: 'The Response of Single Auditory Nerve Fibres to Acoustic Stimulation', *Journal of Neurophysiology*, 1943, **6**, p. 39.
(10) Jeffress, L. A.: 'A Place Theory of Sound Localization', *Journal of Comparative Physiological Psychology*, 1948, **41**, p. 35.
(11) Rosenzweig, M. R.: 'Cortical Correlates of Auditory Localization and of Related Perceptical Phenomena', *ibid.*, 1954.
(12) Firestone, F. A.: 'The Phase Difference and Amplitude Ratio at the Ears Due to a Source of Pure Tone', *Journal of the Acoustical Society of America*, 1930, **2**, p. 260.
(13) Sivian, L. J., and White, S. D.: 'Minimum Audible Sound Fields', *ibid.*, 1933, **4**, p. 288.
(14) Vanderlyn, P. B.: British Patent Application No. 23989, 1954.
(15) De Boer, K.: 'Stereophonic Sound Reproduction', *Philips Technical Review*, 1940, **5**, p. 107.

(16) TERRY, M. T.: British Patent Application No. 12839, 1956.
(17) LIVY, W. H.: British Patent No. 612163, 1946.
(18) MARTIN, M. B., and SMITH, D. L. A.: 'The Design of Magnetic Recording and Reproducing Equipment for Domestic Use', *Journal of the British Institution of Radio Engineers*, 1956, **16**, p. 65.

(9) APPENDIX

(9.1) Change of Position of Apparent Source under Asymmetric Listening Conditions

In the asymmetric case there are three principal reasons for departure from the simple theory applicable to central listening. These are as follows:

(*a*) Operation of the inverse-square law to alter the magnitudes of the sound pressures from the loudspeakers at the listener.
(*b*) The phase difference due to the unequal path lengths from the loudspeakers to the listener.
(*c*) The unequal angles subtended by the loudspeakers at the listener.

The method of calculation of apparent position is shown below. Fig. 13 shows diagrammatically a source S on a sound stage PQ, towards which a microphone pair M is directed. The microphone outputs are reproduced on a similar stage by a pair

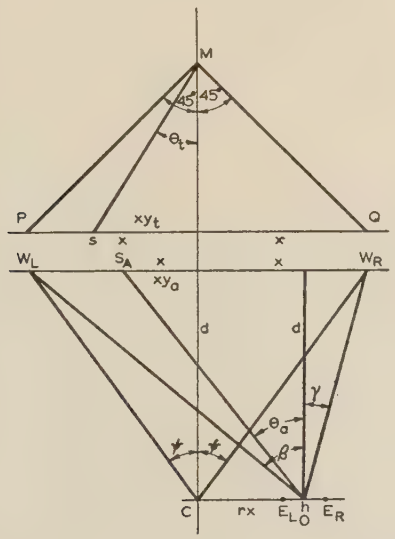


Fig. 13.—Position of apparent source in the asymmetric case.

of loudspeakers, W_L, W_R, assumed to have uniform polar characteristics over the frequency range considered. The virtual image S_a thus produced is observed by a listener O displaced from the central position C.

Then let

$2x$ = Distance between loudspeakers W_L and W_R.
d = Distance between listener O and base-line.
rx = Displacement of listener parallel to base-line.
y_a = Fractional distance of apparent source S from central position.
y_t = Fractional distance of real source from central position.
L, R = Magnitudes of sound pressures due to W_L and W_R at C.
β, γ = Angles subtended at listener by W_L and W_R.
τ = Difference in time of arrival of sounds from W_L and W_R at O due to path difference.
μ_L, μ_R = Time differences at ears E_L and E_R relative to O of sounds from W_L and W_R.
ϕ_L, ϕ_R = Resultant phase relative to O of sound at left and right ears.
θ_a = Apparent angle of arrival of sound at O.
θ_t = True angle of source with respect to microphones for magnitudes L and R from the loudspeakers.

Then:

$$OW_L = d\sqrt{[(1+r)^2\tan^2\psi + 1]}$$

$$OW_R = d\sqrt{[(1-r)^2\tan^2\psi + 1]}$$

$$\tau = \frac{OW_L - OW_R}{v}$$

$$= \frac{d}{v}\{\sqrt{[(1+r)^2\tan^2\psi + 1]} - \sqrt{[(1-r)^2\tan^2\psi + 1]}\}$$

$$\mu_L = \frac{h\sin\beta}{2v}$$

$$\mu_R = \frac{h\sin\gamma}{2v}$$

Relative pressures at the listener's ears are as follows:

	At E_L	At E_R
From W_L	$\dfrac{L\sin\omega(t-\tau+\mu_L)}{\sqrt{[(1+r)^2\tan^2\psi+1]}}$	$\dfrac{L\sin\omega(t-\tau-\mu_L)}{\sqrt{[(1+r)^2\tan^2\psi+1]}}$
From W_R	$\dfrac{R\sin\omega(t-\mu_R)}{\sqrt{[(1-r)^2\tan^2\psi+1]}}$	$\dfrac{R\sin\omega(t+\mu_R)}{\sqrt{[(1-r)^2\tan^2\psi+1]}}$

The resultant phases at the ears relative to O, derived from the above, are:

$$\phi_L = \arctan\left\{\frac{\dfrac{L\sin\omega(\mu_L-\tau)}{\sqrt{[(1+r)^2\tan^2\psi+1]}} - \dfrac{R\sin\omega\mu_R}{\sqrt{[(1-r)^2\tan^2\psi+1]}}}{\dfrac{L\cos\omega(\mu_L-\tau)}{\sqrt{[(1+r)^2\tan^2\psi+1]}} + \dfrac{R\cos\omega\mu_R}{\sqrt{[(1-r)^2\tan^2\psi+1]}}}\right\}$$

$$\phi_R = -\arctan\left\{\frac{\dfrac{L\sin\omega(\mu_L+\tau)}{\sqrt{[(1+r)^2\tan^2\psi+1]}} - \dfrac{R\sin\omega\mu_R}{\sqrt{[(1-r)^2\tan^2\psi+1]}}}{\dfrac{L\cos\omega(\mu_L+\tau)}{\sqrt{[(1+r)^2\tan^2\psi+1]}} + \dfrac{R\cos\omega\mu_R}{\sqrt{[(1-r)^2\tan^2\psi+1]}}}\right\}$$

whence it can be shown that the total resultant phase difference at the ears is:

$$(\phi_L - \phi_R) = \arctan\left\{\frac{\dfrac{L^2\sin 2\omega\mu_L}{(1+r)^2\tan^2\psi+1} - \dfrac{R^2\sin 2\omega\mu_R}{(1-r)^2\tan^2\psi+1} + \dfrac{LR2\sin\omega(\mu_L-\mu_R)\cos\omega\tau}{\sqrt{[(1+r)^2\tan^2\psi+1]}\sqrt{[(1-r)^2\tan^2\psi+1]}}}{\dfrac{L^2\cos 2\omega\mu_L}{(1+r)^2\tan^2\psi+1} + \dfrac{R^2\cos 2\omega\mu_R}{(1-r)^2\tan^2\psi+1} + \dfrac{LR2\cos\omega(\mu_L-\mu_R)\cos\omega\tau}{\sqrt{[(1+r)^2\tan^2\psi+1]}\sqrt{[(1-r)^2\tan^2\psi+1]}}}\right\}$$

Making the approximation, on which the system is based, that $\omega\mu_L$ and $\omega\mu_R$ are small, thus implying also that ϕ_L and ϕ_R are small:

$$(\phi_L - \phi_R) \to \tan(\phi_L - \phi_R) \to \left\{ \frac{\dfrac{L^2 2\omega\mu_L}{(1+r)^2\tan^2\psi + 1} - \dfrac{R^2 2\omega\mu_R}{(1-r)^2\tan^2\psi + 1} + \dfrac{LR2\omega(\mu_L - \mu_R)\cos\omega\tau}{\sqrt{[(1+r)^2\tan^2\psi + 1]}\sqrt{[(1-r)^2\tan^2\psi + 1]}}}{\dfrac{L^2}{(1+r)^2\tan^2\psi + 1} + \dfrac{R^2}{(1-r)^2\tan^2\psi + 1} + \dfrac{LR2\cos\omega\tau}{\sqrt{[(1+r)^2\tan^2\psi + 1]}\sqrt{[(1-r)^2\tan^2\psi + 1]}}} \right\}$$

But $$(\phi_L - \phi_R) = \frac{\omega h \sin\theta_a}{v}$$

Therefore $$\theta_a = \arcsin\left[\frac{v}{\omega h}(\phi_L - \phi_R)\right]$$

Substituting for μ_L and μ_R in the expression for $(\phi_L - \phi_R)$:

$$\theta_a = \arcsin\left\{ \frac{\dfrac{(L/R)\sin\beta}{(1+r)^2\tan^2\psi + 1} - \dfrac{(R/L)\sin\gamma}{(1-r)^2\tan^2\psi + 1} + \dfrac{(\sin\beta - \sin\gamma)\cos\omega\tau}{\sqrt{[(1+r)^2\tan^2\psi + 1]}\sqrt{[(1-r)^2\tan^2\psi + 1]}}}{\dfrac{L/R}{(1+r)^2\tan^2\psi + 1} + \dfrac{R/L}{(1-r)^2\tan^2\psi + 1} + \dfrac{2\cos\omega\tau}{\sqrt{[(1+r)^2\tan^2\psi + 1]}\sqrt{[(1-r)^2\tan^2\psi + 1]}}} \right\}$$

The fractional distance of the apparent source from the central position is then given by

$$y_a = \frac{d\tan\theta_a}{x} - r \quad = \frac{\tan\theta_a}{\tan\psi} - r$$

Similarly, the fractional distance of the real source from the central position before the microphones is given by

$$y_t = \frac{\tan\theta_t}{\tan 45^\circ} = \tan\theta_t$$

45° being half the working angle of the microphones.

DISCUSSION BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 20TH FEBRUARY, 1957

Mr. J. Moir: I am professionally interested in the reproduction of sound films in which three-channel stereophonic reproducer systems are common, but I have an interest in domestic stereophonic systems as an enthusiastic amateur.

I have been playing these stereosonic magnetic-tape recordings for some months, and my only comment on their performance is that the authors have understated the advantages. The spatial distribution and the impression of size that a good stereophonic system gives to an orchestra is most effective and most important. The individual instruments and sections of the orchestra are clearly separated, which gives a degree of clarity and definition to the performance that cannot be obtained from a monaural system, however much the amplitude distortions are reduced. A monaural system with 0·5% distortion does not sound so 'clean' as a stereophonic system with 5% distortion.

It has been appreciated for the last 25 years that the public will not accept a reproduction of an orchestra with the full original frequency range or volume range. Many reasons have been advanced, all with some degree of truth, but there is no doubt that the use of monaural reproducer systems is a major factor in this preference for range restriction.

It has often been said that the advantages of a stereophonic system are too subtle for the ordinary members of the public to appreciate, but my experience tends to the opposite point of view. Fifteen or twenty of my friends have heard these stereophonic tapes run on my domestic equipment and have had the opportunity of making an immediate comparison with a long-playing recording of the same work run monaurally. After about ten seconds of monaural reproduction the invariable comment is on the flatness.

Though I am fairly enthusiastic about these stereophonic recordings I am not so enthusiastic about the authors' explanation of how the system works. They suggest that their two-channel system gives stereophonic results because it allows the brain to compare the amplitude of the signals at the two ears, the signal at each ear being the vector sum of signals reaching the ear from both loudspeakers. They support this suggestion with an elegant mathematical analysis in which I can find no questionable step, but in spite of this, I do not believe that the stereophonic results can be explained in this way or so simply.

A source emitting a single sound pulse in the studio will result in two pulses reaching each ear, but I would suggest that only the first pulse at each ear reaches the brain. The nervous system transmits amplitude indications as a pulse code, pulse-repetition rate being related to sound intensity. However, the maximum pulse rate cannot exceed 1 000 pulses/sec, while 300–500 pulses/sec is the more usual maximum. This pulse-rate limitation is imposed by the maximum rate at which the transmitting-end cells can be chemically recharged by the blood stream. Thus, after a discharge, the nerve cannot transmit a second signal for a time which is always greater than 1 millisec. However, the head dimensions are such that the second acoustic pulse will appear from the remote loudspeaker with a time interval which is always less than 0·6 millisec. At first sight, it would appear that no stereophonic effects could be produced if the sound source emitted short pulses, though in fact the stereophonic effect is most marked.

Our own work indicates that there are clues to source position distributed over the whole of the audio-frequency range though the majority of the information is contained in the frequency range above 500 c/s.

Dr. E. C. Cherry: During the course of a recent informal

meeting* I expressed the opinion that stereophony, using two-loudspeaker systems, is impossible. In view of the authors' demonstration, it seems behoven upon me to explain myself further.

I repeat that it is impossible, but the practice of engineering continually involves the solution of impossible tasks; it is a regular search for compromise. In the present context, the theoretical impossibility arises from several reasons, not the least important being that the acoustic environments in the recording studio and in the home or concert hall are simply not reciprocal. Why, it may be pertinent to ask, did the demonstration take place in the Great Hall of Northampton Polytechnic and not, more conventionally, in the Lecture Theatre? Again, the acoustic conditions may be adjusted to some extent, but they cannot be suitable for all listening situations; neither can the listener be free to move far from a centre-line or other given contour without the apparent directive qualities vanishing.

The binaural directive location of a sound, in real-life environments, is not yet understood; more fundamentally, the psychological 'projection' of sounds to lie 'outside' our heads is not understood, nor are the physical controlling factors. Binaural phase (time) difference and amplitude difference, by themselves, do not account for such projection and angular discrimination.

The brain makes great use of the differences between the signals reaching the two ear portals, and there are several differences other than time and amplitude (e.g. effects of room acoustics and the listener's experience of typical environment effects; head-turning effects; head diffraction; the varied angle of arrival of wavefronts, and impedance mismatch at the ear portals). I agree with Mr. Moir that Fourier-type analysis, and theory such as that given in the paper, are best laid aside when trying to understand stereophonic hearing. Statistical analysis rather than Fourier analysis may be the brain's function.

My remarks do not purport to be disparaging in the least, but I wish merely to stress the need for much more psycho-physical study, not only of binaural hearing, but of other sensory functions in communication. Perhaps then we may understand results such as those which were so well demonstrated by the authors.

Mr. G. Millington: In the early days of electrical recording I did some work under Mr. P. W. Willans on high-fidelity recording of the piano. We found that, with a single microphone in a fixed position, it was extremely difficult to avoid an uneven response in which some notes had a dully wooden sound while others were unnaturally brilliant. When one note was cured by moving the position of the microphone, it was usually at the expense of another in a different part of the register.

I should like to ask the authors whether the placing of the microphone system is similarly critical when using two microphones at the same position but orientated in different directions. As far as I could tell, the recording given was perfect in this respect, and was as fine an example of piano recording as I have heard.

Mr. A. M. Thornton: I assume that the system is intended to have a great future in high-quality and more realistic recording, and that ultimately we shall have this sort of stereophonic quality in our homes.

In the case of the single instrument—the piano—the quality and fidelity were the most satisfying that I have experienced, but when it came to the orchestral item 'Peter and the Wolf' I was very disconcerted. The notes alternated from one side of the stage to the other exactly like the sound of the ping-pong ball in the table-tennis game previously demonstrated. I realized then that that is how the conductor must hear the music, i.e. quite differently from the vast majority of the audience who are sitting some distance away. This raises the interesting point: Does the conductor endeavour to produce a good effect at, say, the middle of the hall, or is he the only person really hearing the music as it should be heard?

In using the system for producing recordings in the home, should the listener be virtually transferred from a position in the auditorium to the conductor's rostrum, or should the aim be to 'seat' him effectively in a position, say, centrally five rows from the front?

Major W. V. G. Fuge: In the demonstration with two people speaking at the same time and recorded stereophonically, it seemed possible to listen at will to either conversation by mentally concentrating without moving the head, so that a mental process could switch the attention from one voice to the other.

Is that the advantage of the stereophonic recording, i.e. that different people listening to the same music can concentrate on any particular instrument they wish from time to time and switch their attention to what pleases them?

Mr. F. Oakes: Mr. Thornton asked whether the stereosonic recording should reproduce what the conductor hears from the rostrum. The answer is in the negative, because the balance of instruments is adjusted during rehearsal to provide the correct sound in the auditorium. Depending on the acoustic properties of the concert hall, this may well mean that what the conductor hears is far from acceptable, whilst correct placement of the microphones further from the orchestra will produce a satisfactory balance, such as would be enjoyed from a good seat in the auditorium.

Mr. J. K. Webb: To anyone who has been privileged to hear a demonstration of the authors' stereosonic reproducing system in an average-size living room, 'revelation' is about the only word adequately to describe it. This scheme is surely a mutation which is bound to set a completely new standard and inevitably expose the deficiencies of 'monosonic' recording.

It must always be borne in mind, when considering domestic sound reproduction, that one cannot contemplate the enormous range of sound levels with which the authors have battered our eardrums in their demonstration, however commendable this may be from the technical angle. Let it be said in praise of music in the home, as it was of Cordelia, 'Her voice was ever soft, gentle and low—an excellent thing in woman'. To circumvent the Fletcher–Munson effect and otherwise wrest the maximum satisfaction from a limited range of loudness is a problem which perhaps merits more attention than it has so far been given.

Mr. R. Vermeulen (*Netherlands: communicated*): There can be no doubt about the benefit, nay the necessity, of stereophony, if one really wants a life-like reproduction and not just 'hi-fi'. By its very nature, even the most perfect loudspeaker cannot possibly be any better than a hole of the same dimensions in the wall of the concert hall. It is still something of a miracle that this hole can be enlarged into a wide window frame by means of only two loudspeakers, but we have to accept this as a fact.

Such a window seat would not be considered quite satisfactory in the concert hall. What we really want is a seat right in the middle of the hall. Therefore, though I agree that at present and for most practical purposes 'the reproduction can be limited to sounds arriving from directions covering an angle of, say, 90° in front of the observer', I do not consider this to be the final solution of perfect music reproduction, but expect a still further development.

It is true that we can imitate the orchestra itself by stereophonic methods, but this imitation orchestra, like the real one, will only sound well in a hall with good acoustics and not in a room with poor acoustics or in the home. In order to have a completely life-like reproduction it is necessary also to simulate the sound waves reflected from the ceiling and from the walls of the concert

* 'The Psychology of Comm unication', *Journal I.E.E.*, 1957, 3 New Series), p. 209.

hall. These are not limited to an angle of 90°, but they come from all sides, and it is essential that they do so. The reverberant sound must not only have the right reverberation time; it must also have the right diffuseness.

It is possible to simulate such a diffuse reverberation artificially by placing loudspeakers all round the auditorium and feeding them with suitably delayed music, using several different time delays. By adding such a stereo-reverberation to the stereophonic reproduction, a completely life-like reproduction can be obtained.

I am impressed by the elegant and convincing way in which the authors have reduced the intensity differences of the loudspeakers to phase differences at the ear of the listener, but I still have some doubt whether binaural hearing can be explained by only one single principle. Phase differences with pure tones are not equivalent to differences in time of arrival of clicks, and I cannot see, at present, how the precedence effect, for instance, can be fitted to the authors' theory. Moreover, there is some experimental evidence that differences of intensity at the ear of the listener cannot altogether be neglected. I am not yet completely convinced that it will be wise to base our techniques on a single effect only.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Mr. H. A. M. Clark, Dr. G. F. Dutton, and **Mr. P. B. Vanderlyn** (*in reply*): As Mr. Moir states, it is agreed that directional clues of various kinds occur at all audible frequencies, and their relative importance must vary to some extent with the spectral content of the sounds with which a system is concerned. In the present case, where the reproduction of music is the chief aim, it is contended that low frequencies play at least as important a part as high frequencies. In the design of a stereophonic system, due regard must be paid to the types of clue capable of reproduction within the physical and economic framework, and if any are found to be mutually exclusive, a choice must be made as to which to reject. In the case of the stereosonic system, it was decided to exclude the 'precedence' clue, since to reproduce it would necessitate specification of the precise listening position to be adopted, from which the observer would not be allowed to move. For this reason, the system will not deal correctly with the case cited by Mr. Moir, i.e. a single short sound pulse. It is fair to say, however, that naturally occurring single sound pulses are infrequent, and analysis of most sounds of short duration reveals them to be composed of many cycles of pressure variation. For localization, the system, as Mr. Moir states, allows a comparison of aural amplitudes, but this occurs at high frequencies only; at low frequencies the mechanism of vector summation described in the paper permits a phase comparison.

Many workers in this field of endeavour will at times be inclined to agree with Dr. Cherry's statement that two-channel stereophony with loudspeakers is impossible. Much depends on the precise meaning attached to the terms 'stereophony' and 'impossible'. He does not deny, however, that reproduced sounds having a pronounced spatial character have been demonstrated, at least to those listeners in the more favoured seats. Perhaps these qualifications may help to illuminate his statement.

In the engineering basis of the system, nothing is assumed about the way in which the human 'black box' operates, nor is any such assumption necessary, since all that is attempted is a reconstruction of the acoustic conditions that would obtain at an observer's ears were he actually present in the recording studio. One requirement for satisfactory reproduction is that the listening room should not add appreciably to the total reverberation, and it should be stressed that the system is designed for domestic use where this is normally the case and where the listener can occupy the optimum position. It is not intended for large-scale demonstration.

Before finally rejecting any explanation, based on Fourier analysis, of the effects produced, it must be remembered that the first operation performed on a sound when it reaches the human cochlea is a Fourier analysis.

In reply to Mr. Millington, the use of a double microphone driving two separate loudspeaker channels does largely avoid the unevenness in tonal quality in some parts of the register caused by standing waves in the studio.

Mr. Thornton dislikes the alternation of sounds between one side of the stage and the other. Although this is certainly what the conductor hears there was no intention to give a reproduced performance from this viewpoint. The arrangement of the demonstration was intended to convey the impression of an orchestra occupying the full area of the stage. It is contended that such an impression would be realistic, but if a domestic user feels that the reproduction is on too large a scale dimensionally, it is always open to him to reduce the distance between his loudspeakers.

Major Fuge is correct in supposing that one of the advantages claimed for the system is an increased freedom for selective listening to a particular part of the recorded sound.

Mr. Webb raises the difficult question of domestic sound levels. It may well be that a sound can be made to appear louder than it really is by controlled juggling with the frequency response of the reproducing channel. The operating range of such a control is likely to be limited, however, and many will think that nothing can recreate the sensation of a really loud sound but reproduction at a level comparable with the original. In this respect he is, to some extent, at the mercy of the recording engineer, since, if modern recordings are not permitted to reach reasonably high levels during reproduction, the quiet passages will be lost in background noise.

Mr. Vermeulen mentions the further application of stereophony to include reverberation from the back of a hall. Many of us who have not had the good fortune to hear the results of his experiments have, nevertheless, read about them with interest, and it may well be that where such means as he describes are applicable and economically feasible, they will add considerably to the subjective enjoyment of a reproduced performance.

Finally, in reply to him and to other speakers, it is not our aim to produce a complete theory of binaural hearing, least of all to ascribe it to a single mechanism. The stereosonic system came about as the result of considering what naturally occurring clues exist, and of attempting to recreate a few of them. Others, such as the 'precedence' clue, have had to be rejected. If the explanations offered to fit the observed facts are not looked upon with favour, it can only be remarked that they are not put forward in any spirit of pontification, but that, at least, they have the merit of some experimental foundation.

621.317.79 : 681.142 : 539.712.4

Paper No. 2337 M
Mar. 1957

AN ANALOGUE COMPUTER FOR NUCLEAR POWER STUDIES

By G. J. R. MacLUSKY, B.Sc.

(*The paper was first received* 10*th July*, 1956, *and in revised form* 7*th January*, 1957. *It was published in March*, 1957, *and was read before a joint meeting of the* Measurement and Control Section *and the* Supply Section 12*th March*, 1957, *held in conjunction with the* British Nuclear Energy Conference.)

SUMMARY

The paper describes a new analogue computer which has been constructed at the Atomic Energy Research Establishment. It is to be used for nuclear power studies and comprises three main sections, namely the reactor-kinetics section, the control-system section and the thermal-system section. The design of these sections and their interconnection is described in detail, and brief reference is made to some of the problems which will be studied with this new computer.

(1) INTRODUCTION

Previous studies of the control of nuclear reactors have discussed the dependence of the reactivity upon the temperatures of various parts of the reactor structure, in addition to the adjustment of the reactivity by control mechanisms.[7, 8] So long as a reactor is to be used only for research, or for plutonium production, it can be regarded as a separate entity for control purposes, and most problems in control and safety can be solved with reference to the reactor neutron kinetics alone. When it is necessary to consider temperature effects, the use of a single-pass cooling system, with an inlet temperature which varies very slowly, allows such consideration to be limited to the reactor structure itself.

However, a nuclear power system will almost always involve the circulation of a primary coolant in a closed loop through the reactor and external heat exchangers, in which heat will be transferred to the working fluid of the turbo-generator. Consequently, the temperatures within the reactor, and its resulting reactivity, will depend on the conditions in the external system. The net reactivity, determined by the combined effect of temperature changes and of control-rod position, will determine the rate of change of mean neutron flux in the reactor, and hence the rate of change of power liberated within it. This power will affect the temperatures within the reactor, the temperature rise of the primary coolant, and hence the external power available. All these thermal interactions will be modified by the thermal capacities of the various structures, and also by the time delays caused by the coolant circulation, and it becomes necessary to consider the complete system as a problem in process control. In addition to the internal thermal feedback described above, there will usually be some external feedback imposed by the movement of the control rod in response to some form of automatic power level-control, and this system will have its own characteristic time-constants.

At an early stage in the feasibility studies of a new reactor it is necessary to study the varied interactions of the complex system outlined above, not only to ensure stability and optimum design of the overall control, but also to evaluate the consequences of a wide variety of possible fault conditions. This is particularly necessary, since a redesign of one component in such a system may involve related changes in many other components.

The large number of independent variables which are involved demands the use of either a digital or analogue computer to give the required results in a reasonable time. Since many of the parameters of the system are only known to a few per cent, there is not appreciable loss of accuracy in using analogue techniques, and the consequent facility for readily changing the parameters of the system and directly observing the resultant behaviour is expected to be particularly useful in the early stages of a reactor design when wide variations in the system are being considered.[1–5] If the analogue computer is made to operate in a real time scale there is the further possibility of replacing sections of it by corresponding real system components, such as control rods and driving amplifiers, to facilitate their functional testing.

In general, the analogue computer will deal only with the variation of various quantities with time. Where there are also spatial variations in the quantities concerned (e.g. neutron flux and fission-product poisoning in different parts of the reactor, temperature distribution across a heat exchanger, etc.), it is normally necessary to carry out a preliminary calculation to determine appropriate average values for the quantities concerned, and these are then used to determine the coefficients for the analogue.[10] Although analogue methods have been applied to the solution of neutron-flux distribution problems, the combination of space and time variations will usually lead to an unduly complex analogue system. However, it is possible in some instances (e.g. xenon poisoning and heat-exchanger temperatures) to approximate to such spatial variations by dividing the structure into a number of zones, and considering the quantity concerned averaged over each zone.

(2) ORGANIZATION OF THE COMPUTER

Two analogue computers[6, 9] have been in use at the Atomic Energy Research Establishment for some time in connection with nuclear reactor studies and for training in reactor operation. The new computer to be described will provide greatly increased facilities for much more comprehensive investigations, particularly in connection with the study of nuclear power problems.

Fig. 1 shows a block diagram of the new computer, its three major sections being for the kinetics of the reactor power, the control system, and the thermal system.

The first section simulates the behaviour of the neutrons in the reactor, and this involves the basic reactor kinetic equations

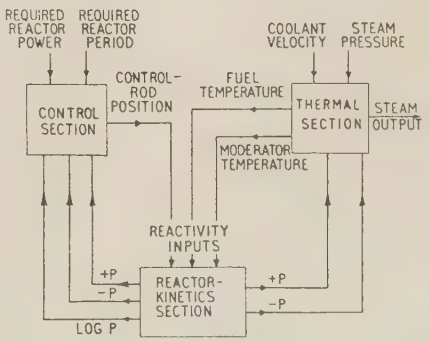


Fig. 1.—Functional arrangement of computer.

Mr. MacLusky is at the U.K.A.E.A. Atomic Energy Research Establishment.

which determine how the rate of change of power level is affected by the net reactivity and by the prompt and delayed neutron lifetimes. The second section simulates the behaviour of the control servo-mechanisms and the electronic amplifiers used to provide voltages proportional to reactor power and its time differential, from which the control programme is derived. The third section simulates the behaviour of the thermal system, involving the temperature changes in the various parts of the reactor (fuel, fuel-element sheath, coolant and moderator), the transfer of heat to the coolant, the time delays in coolant circulation and the transfer of heat through the heat exchanger to produce steam. The system may be extended to give approximate simulation of the turbo-generator and its connection to the load network.

Fig. 1 indicates the quantities which act as the outputs and inputs of each section. It should be noted that all the quantities which can vary the reactivity are supplied separately to the reactor-kinetics section.

A d.c. electronic analogue system is employed, with a number of identical unit amplifiers interconnected by external networks which determine the form of the equations which are solved and their coefficients.

(2.1) Reactor Kinetics

The neutron kinetics can be reduced to a group of linear differential equations, with one additional differential equation which is non-linear, since it involves the multiplication of the reactor power by the net reactivity. Although an electronic multiplier can be used for this operation, it is difficult to achieve sufficient accuracy over the required range in reactor power, and it was decided to use a servo-type multiplier with separate compensation for the finite frequency response of the servo system.

For different reactors the same basic neutron kinetic equations can be used, differing only in the magnitude of various coefficients. Consequently, it is possible to fix the circuit arrangement of this section of the computer and provide a number of controls to vary the required coefficients.

When fully shut down, the neutron flux of a reactor may be 10 decades below the normal operating level, and if an analogue system is to simulate directly a complete reactor start-up, its output voltage must vary over the same range. In one of the existing computers[6] at A.E.R.E. this requirement is met by clamping all voltages after each 10 : 1 excursion in the output, resetting to one-tenth of the values reached, and then continuing the run in a series of such steps.

Since a considerable amount of information on the early stages of reactor start-up is already available, it was decided that the main function of the new computer would be to study reactor operation in the region where thermal effects were of importance. Allowing for a 100 : 1 variation in coolant flow, this means that the analogue must operate continuously over four decades in reactor power, which is possible if vibrator-stabilized d.c. amplifiers are used.

(2.2) Control and Thermal Sections

These sections of the computer have been made very flexible, to allow for very wide variations in possible control and thermal systems in the different reactors to be studied. There is no fixed circuit arrangement, but 40 unit amplifiers and 20 coefficient units of each of two types are available. These are interconnected by coaxial cables to set up the required equations. In addition, eight servo-multiplier units can be used to give coefficients which vary with time, or as a function of chosen variables in the equations. For instance, the rates of heat transfer from fuel to coolant, or from coolant to heat exchanger, can be made to depend on the coolant velocity as well as the temperature differences.

(2.3) Coolant Delays

Pure time lags may be produced by the flow of coolant through long pipe runs without appreciable mixing. In terms of electrical networks these will be equivalent to delay lines having zero attenuation and a phase change proportional to frequency. When mixing occurs, giving additional high-frequency attenuation, a rough approximation to the resultant characteristic can be obtained by a cascaded series of *RC* networks.

However, it is desirable to have a more accurate simulation of such time delays, in order to use the analogue to determine the amount of mixing necessary for thermal-system stability. Moreover, the time delay must depend on the coolant velocity, which may itself vary with time. Equipment is under development in which the quantity which must be delayed (e.g. coolant temperature) is recorded on a loop of magnetic tape whose speed can vary in proportion to coolant velocity. A play-back head suitably positioned on the same tape loop will then give the required delayed signal.

(2.4) Layout of the Equipment

In the complete computer, the first four cabinets house the thermal section, the fifth houses the two 3-pen recorders, the sixth the reactor-kinetics section and the seventh the control section. In all cabinets except the fifth and sixth the top four units consist of mounting panels into each of which three unit amplifiers and three coefficient units can be mounted, giving a total of 12 amplifiers and 12 coefficient units per cabinet. The next three units in the cabinet are a duplex servo-multiplier unit and the two servo amplifiers associated with it.

In the sixth cabinet the top mounting panel carries four linear amplifiers and one logarithmic amplifier associated with the reactor-kinetics section. Below this are the kinetics-components section, the reactivity section (which also includes one servo multiplier) and a control panel associated with the overload circuits.

The control desk, which is carried by the fifth cabinet, has output sockets from all the amplifier positions in the computer and initial-condition input sockets for all coefficient units. Recorder, meter, and overload-detecting-circuit input sockets also appear on the control desk, so that, with the aid of short wander-leads, any amplifier output can be read on the meter, or by one of the six recorder channels, while initial conditions can be set by the 24 potentiometers provided for this purpose.

When the magnetic-tape delay unit is completed it will be housed in an eighth cabinet.

(3) BASIC UNITS

(3.1) Unit Amplifier

The unit amplifier consists of a 3-valve d.c. amplifier with a vibrator drift-correcting system employing a further two valves (Fig. 2). The vibrator used has very good shielding between the 50 c/s driving coil and the contacts, resulting in low spurious currents at the input. The d.c. amplifier has a response which is flat to approximately 130 c/s; the high-frequency gain is controlled by the usual *RC* networks to fall at 9 dB/octave. The amplifier characteristics are given below.

	With drift correction	*Without drift correction*
Gain	8×10^6 at d.c. 4×10^3 from 1 to 130 c/s	4×10^3 from 0 to 130 c/s
Drift	$\pm 50\,\mu$V	± 5 mV, short-term ± 50 mV, long-term
Current at input ..	$< 5 \times 10^{-11}$ amp	$< 5 \times 10^{-12}$ amp

The front panel dimensions allow six such units to be assembled into a cut-out in a standard 7 in × 19 in panel.

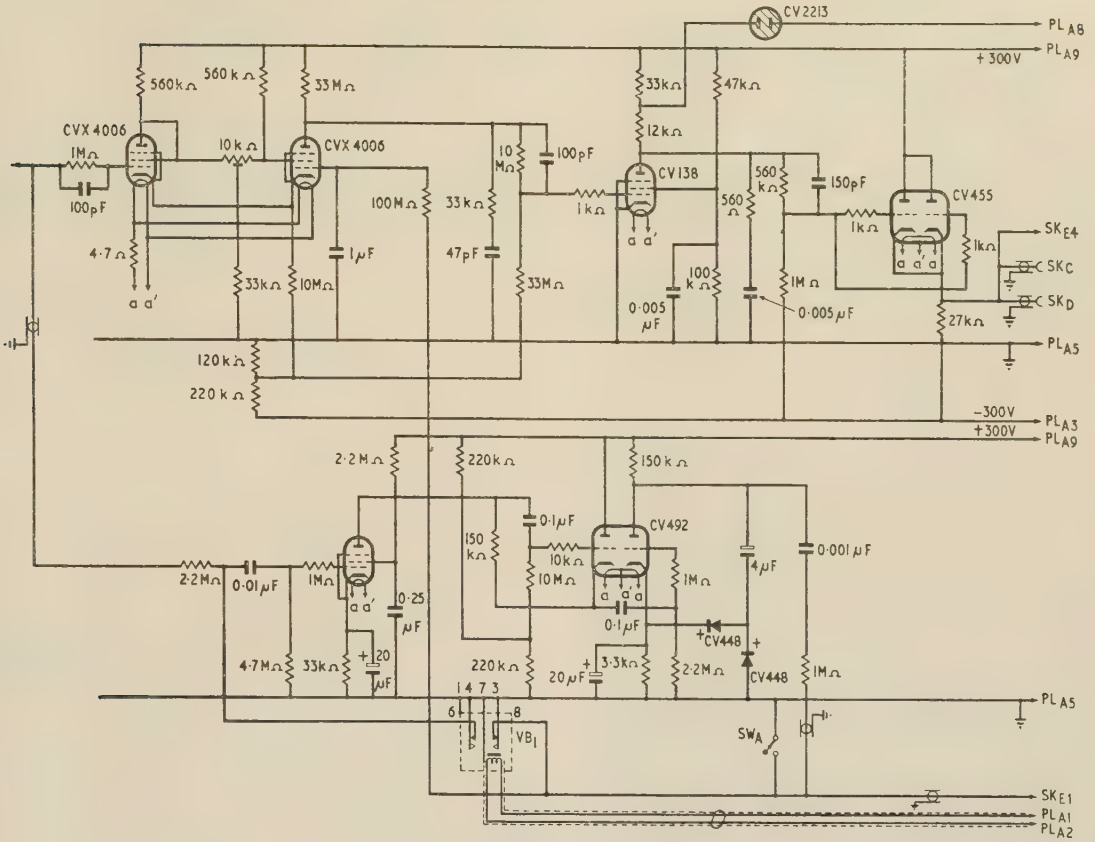


Fig. 2.—Circuit of unit amplifier.

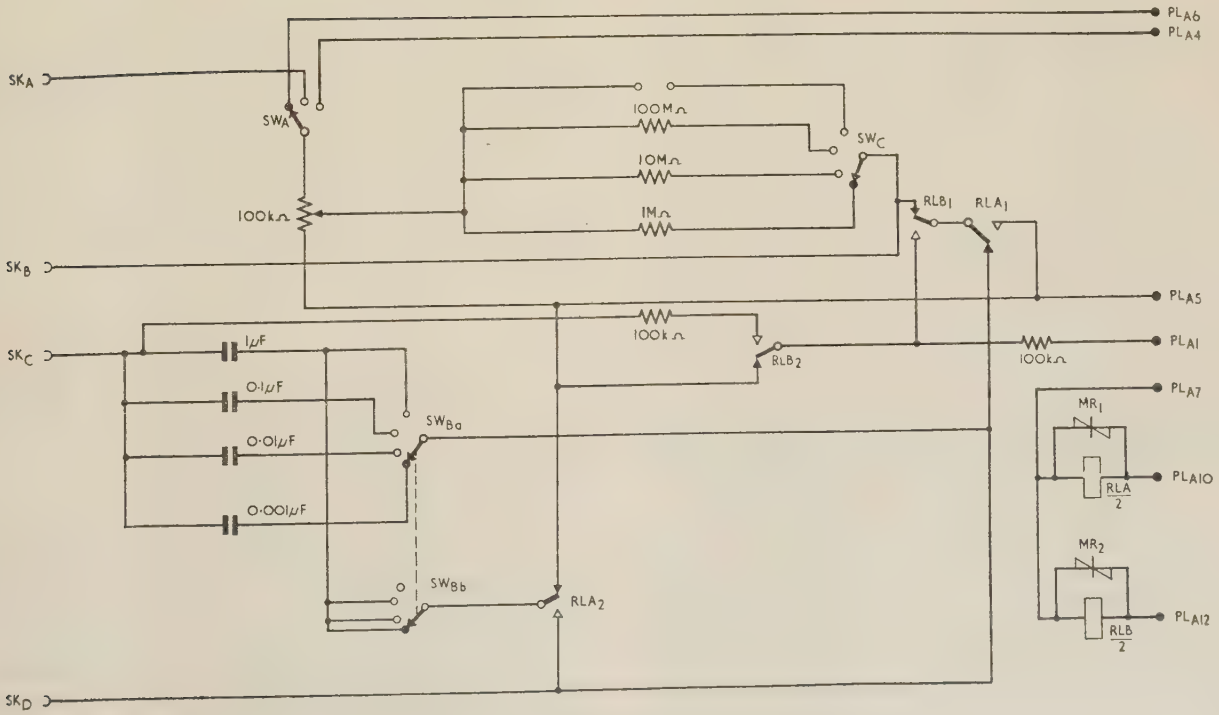


Fig. 3.—Circuit of *RC* coefficient unit.

(3.2) Logarithmic Amplifier

One logarithmic amplifier is used in the reactor-kinetics section, being fed with a current proportional to the reactor power. It was developed by Wilson and Cox,[9] and uses the control grid of a pentode as the anode of a logarithmic diode, with negative feedback applied to the screen grid of the same valve. In addition, a vibrator drift corrector adjusts the cathode potential of the pentode to hold the grid potential to within 50 μV of earth, and the output voltage from the screen is accurately logarithmic over five decades, from 50 volts to 500 μV at the input.

Differentiation of this output will then give the reactor period, which is proportional to the inverse of doubling time. In simulation of an automatic start-up system the period signal is used as the major control quantity.

(3.3) Coefficient Units

The two types of coefficient unit have the same front-panel dimensions as the amplifier. The first type (Fig. 3) provides a capacitance (used as a feedback element) variable in 10 : 1 steps from 0·001 to 1 μF. The 100-kilohm helical potentiometer which feeds a resistor switched to 1, 10 or 100 megohms, provides a transconductance which is continuously adjustable over a range of more than 10^4 : 1. The fourth switch position is left blank to allow for the addition of a high-value resistor to cover the few instances where a time-constant longer than 100 sec is required.

In normal operation both A- and B-relays (Fig. 3) will be released, with their contacts as shown. Operation of A-relays will clamp all integrators at the voltage they had reached, and since the 1 μF capacitor is then connected across the amplifier the drift due to residual currents will be only 50 μV/sec. Operation of B-relays will enable initial conditions to be set by a suitable voltage applied to PL_{A1}.

Fig. 4 shows the circuit of the second type of coefficient unit,

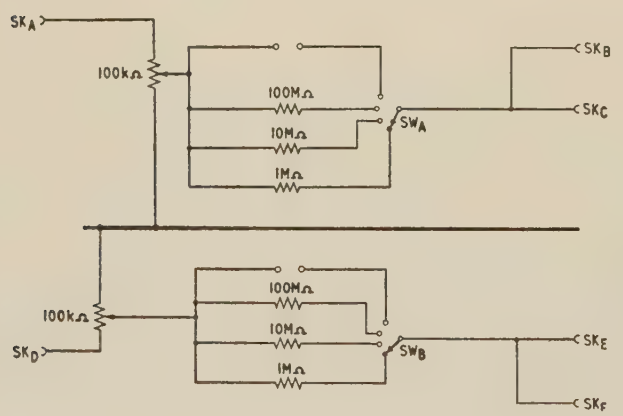


Fig. 4.—Circuit of double-R coefficient unit.

which consists of two identical resistance sections. Each has two output connectors, allowing for the connection of any number of such units in parallel if required by the problem.

(3.4) Servo-Multiplier Unit

Two independent servo systems are mounted in each unit. Each has two potentiometers driven from an a.c. servo motor through a double 60 : 1 reduction gear. By means of additional couplings, up to eight further potentiometers can be added. Each potentiometer has a winding accuracy of 0·1% and is provided with four intermediate tapping points so that non-linear laws can be produced by adding external loading resistors. This is expected to be particularly useful when it is required to multiply various quantities by different non-linear functions of one other quantity; e.g. heat-transfer rates as a function of coolant velocity.

(4) SYSTEM SIMULATION

The methods by which the basic units are interconnected to solve the required differential equations are the same as for other d.c. analogue computers. For instance, Fig. 5 shows how one

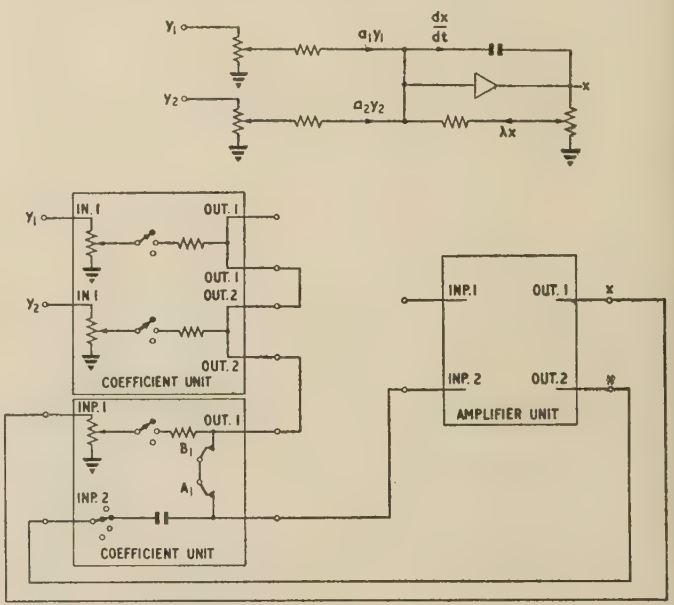


Fig. 5.—Analogue system to solve $dx/dt = a_1 y_1 + a_2 y_2 - \lambda x$.

unit amplifier and two coefficient units would be applied to solve an equation of the type

$$dx/dt = a_1 y_1 + a_2 y_2 - \lambda x \quad . \quad . \quad . \quad . \quad (1)$$

By adding a second integrator, an equation involving d^2x/dt^2 could be solved, such as the response of a control-rod servo-mechanism.

(4.1) Thermal System

As an illustration of the type of equation encountered in the thermal system, the average temperature of a certain length of uranium fuel element in a reactor is given by

$$\rho_u S_u C_u \frac{d\bar{T}_u}{dt} = S_u a P - ph(\bar{T}_u - \bar{T}_c) \quad . \quad . \quad . \quad (2)$$

where P = Reactor power.
a = A constant, depending on the heat developed per unit length of fuel.
S_u = Cross-sectional area of uranium fuel.
ρ_u = Density of uranium fuel.
C_u = Specific heat of uranium fuel.
$\bar{T}_u$ = Temperature of uranium fuel, averaged over the length concerned.
$\bar{T}_c$ = Temperature of coolant, averaged over the length concerned.
p = Perimeter of fuel in contact with coolant.
h = Heat-transfer coefficient.

For a gas or normal liquid,

$$h = a_1 v^{0\cdot 8} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3a)$$

For a liquid metal,

$$h = a_2 + a_3 v^{0\cdot 8} \quad . \quad . \quad . \quad . \quad . \quad (3b)$$

where v is the coolant velocity and a_1, a_2, a_3 are constants.

To solve eqn. (2) with constant coolant velocity, only one unit amplifier and associated coefficient units are required, as for eqn. (1). For varying coolant velocity it is necessary in addition to use a servo system to generate the term varying as $v^{0\cdot 8}$. The temperatures and rates of heat transfer between other parts of the reactor structure, and between the primary coolant and the working fluid in the heat exchangers, can be dealt with by extensions of the above methods.

(4.2) Reactor-Kinetics Section

The equations governing the neutron kinetics of a nuclear reactor can be written as

$$l^* \frac{dn}{dt} = (1 - \beta)Kn - n + \sum_{i=1}^{i=12} \lambda_i x_i + S \quad . \quad . \quad (4)$$

$$\frac{dx_1}{dt} = -\lambda_i x_i + \beta_i Kn \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where l^* = Mean lifetime of neutrons from generation to absorption.
n = Neutron concentration.
K = Effective multiplication factor.
x_i = ith delayed neutron emitter concentration times l^*.
λ_i = Decay rate of ith delayed-neutron emitter.
S = Neutron source in neutrons per l^* seconds.
β_i = Total fraction of neutron production which is delayed by the ith group.
β = Total fraction of neutron production which is delayed by all delayed groups.
(By definition $\sum_{i=1}^{i=12} \beta_i = \beta$).

The derivation of these equations is given in Section 7.1. In general there may be up to 12 equations of type (5).

For most purposes the reactor power output can be taken as being proportional to the neutron flux, and the relatively small power due to heating by the decay of fission products can be computed separately if necessary. Furthermore, the multiplication factor is normally very close to unity, and the above equations can be simplified, giving

$$l^* \frac{dP}{dt} = (\delta - \beta)P + \sum_{i=1}^{i=12} \lambda_i x_i + S \quad . \quad . \quad . \quad (6)$$

and

$$\frac{dx_i}{dt} = -\lambda_i x_i + \beta_i P \quad . \quad . \quad . \quad . \quad . \quad (7)$$

treating

$$\left.\begin{aligned} \beta KP &\simeq \beta P \\ \beta_i KP &\simeq \beta_i P \end{aligned}\right\} \text{when } K \simeq 1$$

where P = Reactor power, and $(K - 1) = \delta$ = 'reactivity'. In practical instances, usually, $0\cdot 95 < K < 1\cdot 01$. This simplification reduces the original 13 non-linear equations to 12 linear and one non-linear equations, in which form an analogue solution can more easily be obtained.

The analogue system used to solve these equations is shown in Fig. 6, in which for clarity only the first two of the 11 delayed-neutron networks are shown. The relay contacts A–L, Q and T are concerned with clamping and establishing initial values in this section of the computer. On the diagram they are all shown in the unoperated condition, but in the 'run' condition all except T are operated. In Section 7.2 the equations for this network

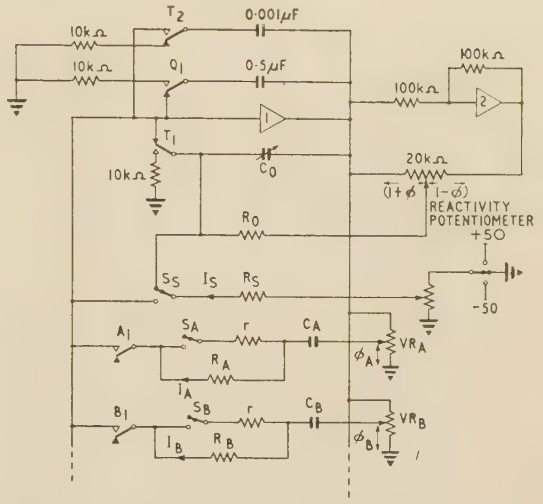


Fig. 6.—Analogue system for reactor-kinetics section.

are derived, and it is shown that they correspond to the simplified reactor eqns. (6) and (7).

Although only six delayed-neutron circuits would be necessary to represent a reactor having a graphite moderator, in reactors using heavy water as moderator the γ-ray-emitting fission products induce corresponding neutron emission from the deuterium, producing six more significant delayed-neutron sources. The last of these, having a mean life of 9 640 sec, has been omitted, but circuits are provided for the 11 other groups, of mean lifetimes ranging from 0·07 to 2 340 sec, and their abundances are set by the associated potentiometers VR_A, VR_B, etc. The prompt-neutron mean lifetime is determined by C_0, and can be varied from 5 millisec to 1 microsec, to cover the range of thermal, intermediate and fast reactors. With switch S set as shown this section will operate in 'real' time. In its second position all the time-constants R_AC_A, etc., are reduced to fractions of a second, and this is used when setting initial conditions. In a third position of switch S (not shown in the diagram) all the delayed-neutron resistances R_A–R_L are reduced to exactly one-tenth of their former values, giving operation in a time-scale accelerated by a factor of ten. R_0 is provided with a switch control to give reactivity ranges of $\pm 10^{-3}$, $\pm 10^{-2}$, and $+10^{-2}$ to -4×10^{-2} in real time, or $\pm 10^{-4}$, $\pm 10^{-3}$, and $+10^{-3}$ to -4×10^{-3} in accelerated time.

The reactor-kinetics section of the computer will operate accurately over four decades of reactor power, which is expected to be adequate for the majority of investigations. To meet the occasional requirement for a wider range, provision is also made to clamp all voltages, to meter them separately, and then to reset all to some suitable fraction of the previous value before continuing the run. In the clamped condition only relay T is operated. When one of the relays A–L is then operated, accompanied by the operation of Q, amplifier 1 is used as an infinite-impedance voltmeter to measure the charge on the corresponding delayed-neutron capacitor, and a similar routine allows all voltages to be individually reset.

As shown in Fig. 7, an a.c. servo-motor is used to position the reactivity potentiometer in response to the total reactivity input current, which will normally be derived from a number of coefficient units in the other sections of the computer, fed by the various quantities which affect the reactivity. Switch S_1, which determines the reactivity scale, is also arranged to cancel non-linearities by equalizing the loading on both sections of the ganged potentiometer. If the servo system had instantaneous response the input point would be maintained as a virtual earth,

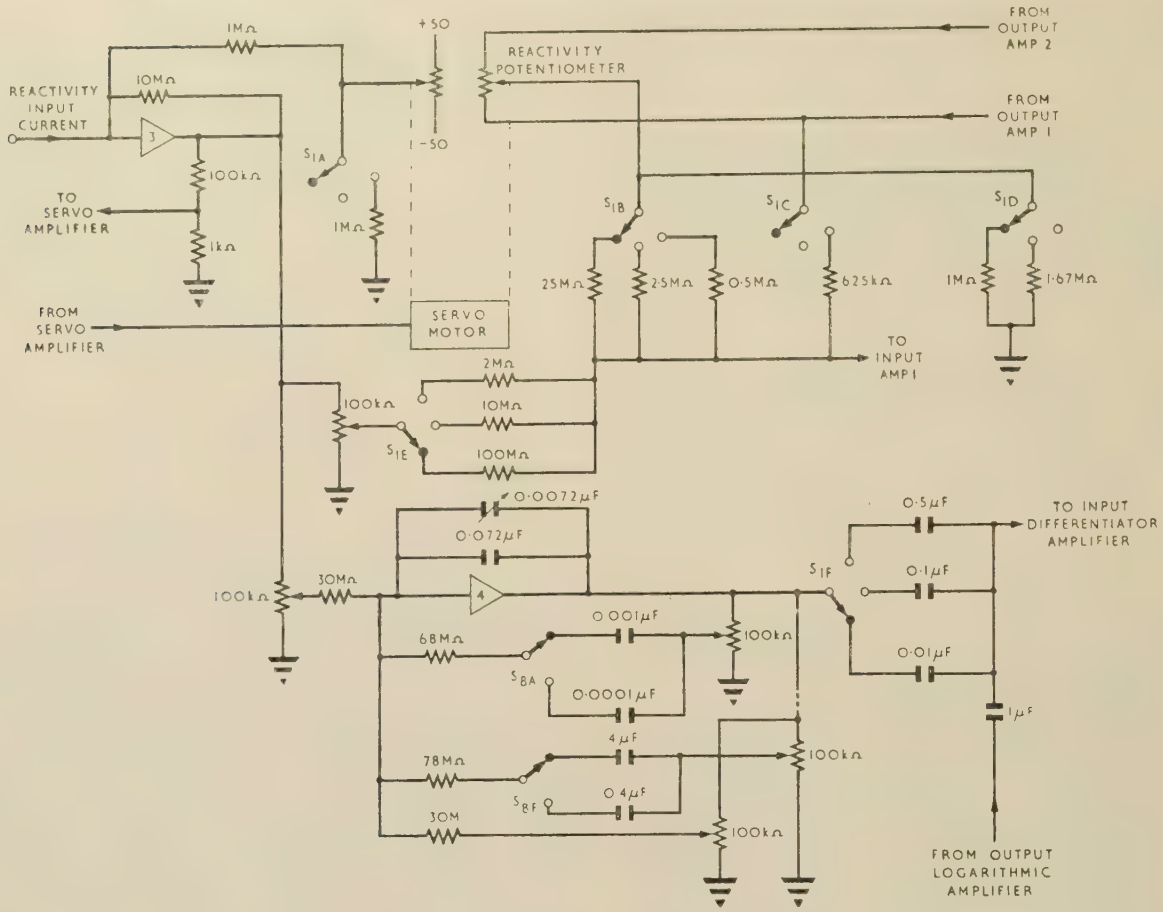


Fig. 7.—Arrangement of reactivity section.

but in practice the finite acceleration and frequency response of the servo will cause an error signal at that point which is amplified by amplifier 3. (The subsequent 100 : 1 degeneration is necessary to make the servo amplifier interchangeable with these in other positions in the analogue.) The constants are so chosen that ± 50 volts on the output of amplifier 3 will correspond to a $\pm 5\%$ full-scale position error in the servo potentiometer, i.e. a reactivity error of $\pm 5\%$ of the full-scale reactivity change.

This error voltage is then used to compensate for the finite servo response in two alternative ways. In the first, a current proportional to the servo error, but scaled according to the position of switch S_1, is fed to the reactivity input of the reactor-kinetics section, replacing the current which was 'missing' because of the finite servo response. Since this current should also be proportional to the voltage across the reactivity potentiometer (i.e. the reactor power at that instant), the 100-kilohm potentiometer at the input to amplifier 4 is used to adjust the amount of compensation in proportion to the output voltages of amplifiers 1 and 2. Clearly this form of compensation is most applicable to studying small transients in reactor power on either side of a fixed power, as may occur when analysing automatic power-level-control systems.

In the second method, the same servo error voltage determines the input current to the feedback circuits associated with amplifier 4, which are arranged to match the prompt- and delayed-neutron circuits in the reactor-kinetics section. However, it is necessary to include only those delayed-neutron groups whose mean lifetime is comparable with the likely duration of the servo transients, although it must be remembered that when operating in an accelerated time-scale these transients are effectively lengthened by a factor of 10 by comparison with the delayed-neutron mean lifetimes. Therefore only groups up to F ($T_0 = 31{\cdot}2$ sec) are included, the remainder being represented by the 30-megohm feedback resistor.

The output of amplifier 4 will now represent a signal which, if added with appropriate scale factor to the output of amplifier 2, would give the correct output voltage in the absence of servo errors. But the output of amplifier 2 also feeds a logarithmic amplifier, which for small input transients will have a gain inversely proportional to the mean value of the input. In other words, a given fractional change at the input to the logarithmic amplifier will always result in the same absolute change at the output, independent of the actual magnitude of the input. Consequently, if the output from amplifier 4 were added to the logarithmic output, the only scale adjustment needed would be to allow for the different position of S_1. Since the logarithmic output normally feeds a differentiator circuit, it is more convenient to add the correction at its input, and S_{1F} switches the differentiator input capacitor appropriately. The result of this method of compensation will be that the output voltages corresponding to the reactor power will not be compensated for finite servo response, but that the reactor-period signal obtained from the differentiator will be accurately compensated, even though the reactor power varies over many decades. Consequently, this method of compensation is applicable to studying the automatic starting-up of a nuclear reactor, where reactor period is an important control quantity.

Both these methods of compensation are equivalent to replacing the non-linear equation (6) by a linear approximation which is valid only if the required correction to P is a small fraction of its

value at the start of the transient concerned. In practice this is true, since it is unlikely that the reactivity input will combine high rate of change and large amplitude, and the reactor itself has a response which falls with increasing frequency.

In addition to the servo-controlled reactivity input, truly instantaneous reactivity changes can be produced by switching fixed resistors from the outputs of amplifiers 1 or 2 to the input of amplifier 1. Alternatively, the same amplitudes of reactivity change can be caused by injecting proportional current transients into the reactivity input of the servo system, and for test purposes these two effects can be combined in opposition. This facilitates adjustment of the compensation for transient response of the servo mechanism, since the remaining transient on the output is produced by the discrepancy between the true step function in reactivity and the same transient distorted by the servo response.

(5) OPERATION

(5.1) Servo Compensation System

Fig. 8 shows the results obtained from tests of the servo compensation system and the reactor-kinetics section. Response (*a*) shows the form of reactivity input used throughout the tests, the extreme amplitude being $\pm 5 \times 10^{-4}$ in reactivity. Response (*b*) is the actual servo error resulting, i.e. the output voltage of amplifier 3, and it is seen that the servo has a time-constant of approximately $\frac{1}{5}$ sec. Curve (*c*) is the response of the reactor-output-power voltage to the true reactivity transient, not distorted by the servo error. The delayed-neutron groups used were of mean lifetimes 0·07, 0·62, 2·19, 6·51, 31·7 and 80·2 sec and the prompt-neutron mean lifetime was 10^{-4} sec. Curve (*d*) is the response of the reactor output to the same transient distorted by the servo, while curve (*e*) is the result of true and distorted reactivity transients in opposition. For both (*d*) and (*e*) the compensation was zero, while the result shown in (*f*) was obtained as in (*e*) but with optimum compensation. Response (*g*) was obtained as for (*d*), but with optimum compensation, and should be compared with (*c*). Curve (*h*) is the response of the recorder itself to a square-wave input of effectively zero rise-time. The rise time of the recorder is seen to be 20 millisec substantially shorter than those of the waveforms recorded.

When a test similar to (*c*) is carried out with the prompt-neutron mean lifetime increased to 1 millisec, the initial rate of change of reactor output is slower, being about 160 millisec for 0·65% change in output, compared with 60 millisec for a prompt-neutron mean lifetime of 0·1 millisec. This is to be expected, since the initial fractional rate of change of reactor power is given by δ'/l^*, where δ' is the change in reactivity and l^* is the mean lifetime of prompt neutrons. However, this means that for larger mean lifetimes the deterioration in the observed response due to the servo time-lag is much less serious, and there is less apparent improvement with the servo compensation.

Similar tests have been made on the second method of servo-error compensation, which involves the addition of the outputs of the logarithmic amplifier and of amplifier 4 which generates the compensation signal. At the outset the feedback networks across amplifiers 1 and 4 must be matched in response, and since the frequency range over which the response is affected by any one *RC* network can be calculated, it proved convenient to observe the frequency response rather than the transient response. This was done by feeding an appropriate small alternating current into the input point of the amplifier under test, and by successive adjustments of the feedback potentiometers associated with amplifier 4 the two characteristics were matched to within 1 dB over the range 0·01–100 c/s. (This adjustment should not need to be repeated, since any change in the proportions of delayed neutrons in the reactor-kinetics section will involve only proportional changes in the compensation system.)

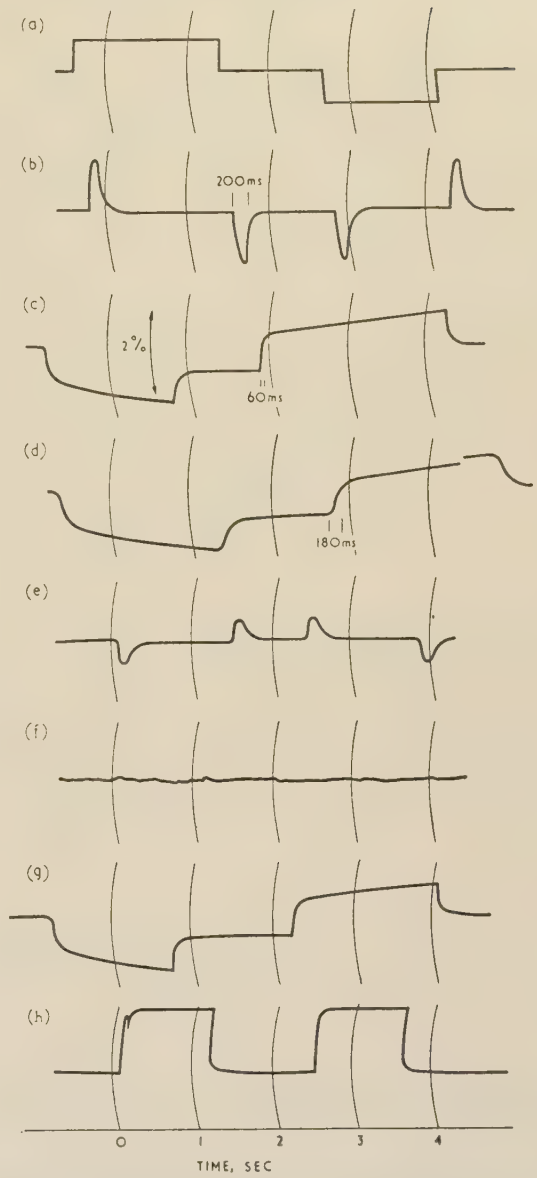


Fig. 8.—Transient responses in reactor-kinetics section.
Since the timing of the edges of the input (*a*) was variable, the time displacements of the successive outputs (*b*)–(*g*) are not significant.

The overall response of the servo system and reactor-kinetics section was observed by feeding the test alternating current into the reactivity input and measuring the combined output of logarithmic and compensation amplifiers. With the amplitude of compensation reduced to zero, the measured response began to deviate appreciably from the true response above 0·5 c/s, and was down 30 dB at 100 c/s. With the optimum amplitude of compensation the discrepancy was reduced to less than 2 dB over the frequency range 0·01–100 c/s.

(5.2) Thermal System

It will be useful to illustrate the use of the thermal section by referring to the use of the analogue computer to simulate the operation of an automatic control system in the presence of appreciable temperature coefficients of reactivity due to fuel

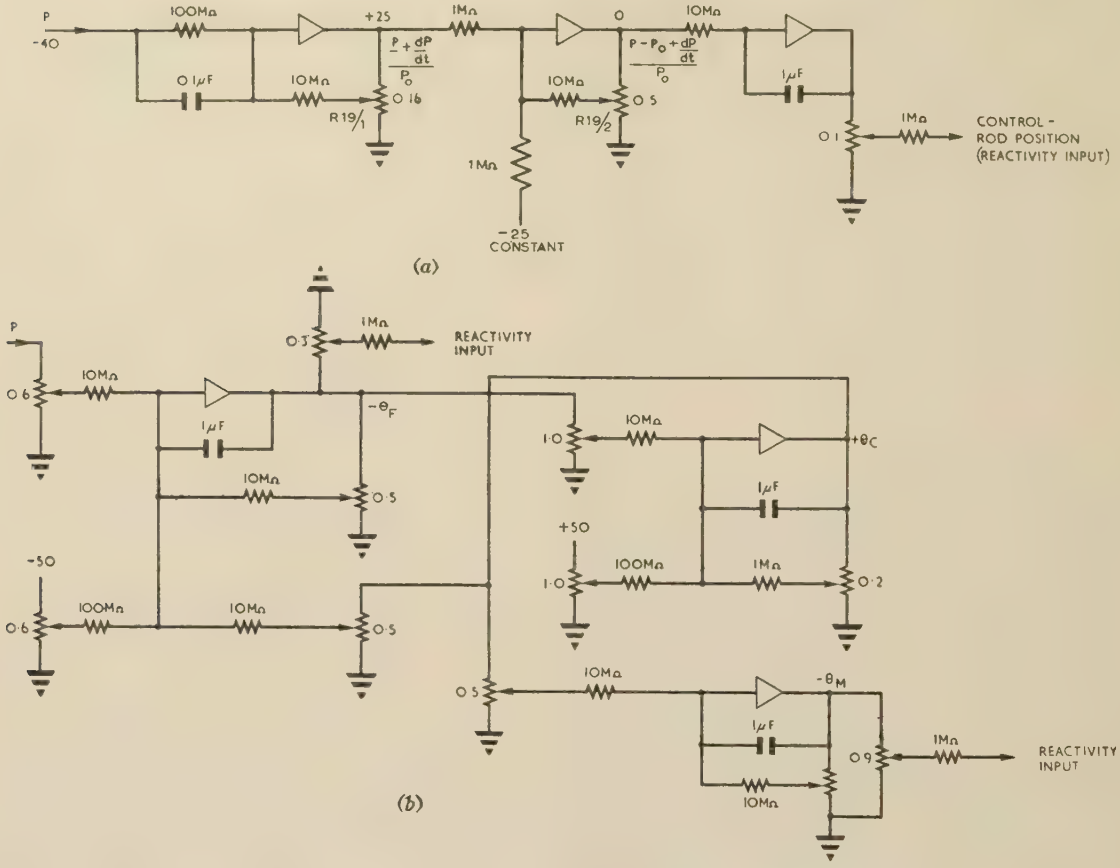


Fig. 9.—Analogues.

(*a*) Automatic power-level control. (*b*) Reactor thermal system.
θ_F = Fuel temperature. θ_C = Coolant temperature. θ_M = Moderator temperature.

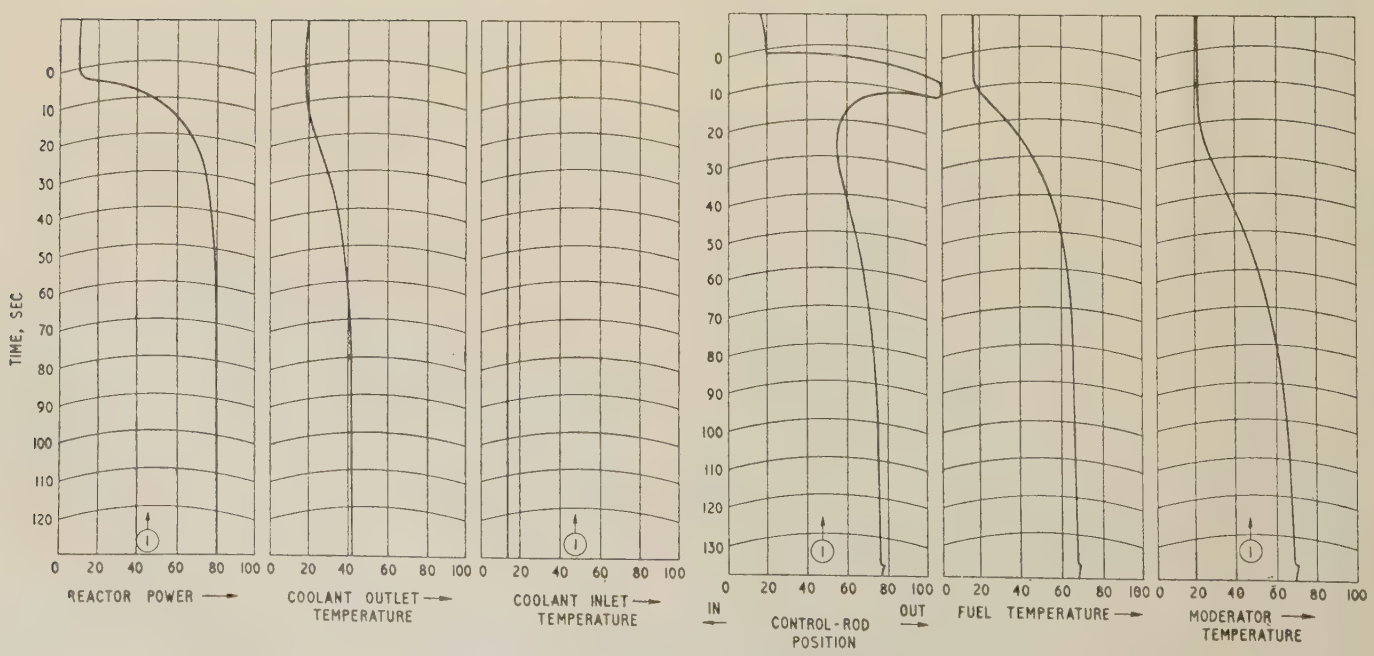


Fig. 10.—Response of reactor and thermal system.

and moderator temperature changes. The system used is shown in Fig. 9. Fig. 9(*a*) shows the arrangement of the automatic control section, R19/1 being the control for the 'demanded' power. Fig. 9(*b*) shows the corresponding arrangement of the thermal system of the reactor with constant inlet coolant temperature.* The numbers opposite each potentiometer give the fractional rotation to which it was set during the tests.

Initially R19/1 was set at 0·016, corresponding to a demanded power of 4 units from the reactor-kinetics section. Then R19/1 was rapidly reset to 0·16, corresponding to a demand for 40 units. Fig. 10 is a reproduction of the recorder charts, and shows the initial rapid movement of the control rod to produce the required rate of increase of power. However the 0·1 μF capacitance at the input to amplifier 2 means that the output voltage of that amplifier depends on the rate of change of power as well as the difference between actual and demanded powers. As the power approaches the new demanded level the required rate of change decreases, and to satisfy this the control rod is moved towards its original position, which it would reach in the absence of thermal effects. However, the relatively rapid increase in fuel temperature followed by the slower increase in moderator temperature cause a decrease in net reactivity. To maintain the new power in spite of these changes, the control rod is made to move out again and will not reach a new steady position until the moderator temperature has reached equilibrium.

(5.3) Operating Experience

Among the problems which have been studied with the computer are investigations of the transients in reactor period resulting from various rates of linear reactivity change and in reactor power and fuel temperature following a change of fuel concentration in a homogeneous reactor. The facility for readily monitoring any point in the system has proved of great value when setting up a new problem on the computer, and it is apparent that the time required for this operation decreases as experience is gained on this particular equipment. To ensure efficient utilization of a computer of this type, it would seem desirable that one member of the staff should have prime responsibility for translating the variety of problems presented into a suitable analogue form and setting them on the computer.

(6) REFERENCES

(1) KORN, G. A., and KORN, T. M.: 'Electronic Analog Computers' (McGraw-Hill, New York, 1952).

(2) SCHULTZ, M. A.: 'Control of Nuclear Reactors and Power Plants' (McGraw-Hill, New York, 1955).

(3) GLASSTONE, S., and EDLUND, M. C.: 'Elements of Nuclear Reactor Theory' (McGraw-Hill, New York, 1955).

(4) BELL, P. R., and STRAUSS, H. A.: 'Electronic Pile Simulator', *Review of Scientific Instruments*, 1950, **21**, p. 760.

(5) PAGELS, W.: 'A Portable Electronic Pile Kinetic Simulator', A.I.E.E. Technical Paper, 51–262, May, 1951.

(6) WILSON, I., and BURNUP, T. E.: 'An Analogue Computer for Reactor Kinetic Equations', A.E.R.E. Report RE/R.1485.

(7) COX, R. J., and WALKER, J.: 'The Control of Nuclear Reactors', *Proceedings I.E.E.*, Paper No. 2068M, March, 1956 (**103** B, p. 577).

(8) BOWEN, J. H.: 'Automatic Control Characteristics of Thermal Neutron Reactors', *ibid.*, Paper No. 1432M, December, 1952 (**100**, Part I, p. 102).

(9) WILSON, I.: 'A Reactor Training Simulator', A.E.R.E. Report R/R.1917.

(10) STONE, J. J., and MANN, E. R.: 'Oak Ridge National Laboratory Reactor Controls Computer', ORNL-1632.

* The parameters employed do not correspond to any specific reactor.

(7) APPENDIX

(7.1) Derivation of Reactor Kinetic Equations

A nuclear reactor consists of an assembly of fissionable material, such as uranium 235 (possibly associated with non-fissionable material, such as uranium 238, and a moderator such as graphite) of such dimensions that a fission chain reaction can be maintained. If at any instant there are in a unit volume of the reactor n neutrons which have just been produced by fission, these will have a mean lifetime of l^* before they are absorbed, escape from the reactor, or collide with new atoms of fissionable material to produce nK new neutrons. The constant K, which gives the ratio of successive neutron generations, is termed the effective multiplication constant for the particular reactor, and must be equal to or greater than unity for the chain reaction to proceed.

If it is assumed that all the neutrons from fission appear instantaneously, the rate of change of the neutron population can be expressed by the equation

$$\frac{dn}{dt} = \frac{Kn - n}{l^*} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In uranium 235, however, about 0·75% of the fissions give rise to a series of groups of delayed neutrons (the proportion is different in other fissionable materials). If allowance is made for photo-neutron production, there may be up to 12 such groups, each having a characteristic radioactive decay constant λ. If C_i is the concentration of the sources of delayed neutrons in the ith group, the rate of production of such sources will be a known fraction β_i of the total rate of production of new neutrons in fission, and we can write

$$\frac{dC_i}{dt} = \frac{\beta_i K_n}{l^*} - \lambda_i C_i \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Multiplying by l^*, and substituting $X_i = l^* C_i$, we have

$$\frac{dX_i}{dt} = \beta_i Kn - \lambda_i X_i$$

which is eqn. (5).

The total production of delayed-neutron sources reduces the rate of production of prompt neutrons by an amount $\frac{\beta Kn}{l^*}$, where $\beta = \sum_{i=1}^{i=12} \beta_i$. On the other hand, the delayed neutrons, as they appear, must contribute an increase of total neutron population at a rate $\sum_{i=1}^{i=12} \lambda_i C_i$. Making these changes to eqn. (8), and also allowing for a constant source of neutrons (e.g. from spontaneous fission) of intensity S per l^* seconds, gives

$$\frac{dn}{dt} = \frac{Kn - n}{l^*} - \frac{\beta Kn}{l^*} + \sum_{i=1}^{i=12} \lambda_i C_i + \frac{S}{l^*} \quad . \quad . \quad (10)$$

Multiplying by l^* and rearranging we have

$$l^* \frac{dn}{dt} = (1 - \beta)Kn - n + \sum_{i=1}^{i=12} \lambda_i X_i + S$$

which is eqn. (4).

(7.2) Analogue System for Reactor Kinetics

To derive the equations of the system shown in Fig. 6, let the current flowing in the ith delayed-neutron network be I_i and the charge on the corresponding capacitor be Q_i. It may be assumed

that the input of amplifier 1 is at earth potential. Then, considering only C_0, gives

$$C_0\frac{dV}{dt} = \frac{\phi V}{R_0} + \sum_{i=1}^{i=12} I_i + I_s \quad . \quad . \quad . \quad . \quad (11)$$

where ϕ is the fractional displacement of the reactivity potentiometer towards the end connected to the output of amplifier 2.

Consideration of only the ith delayed-neutron network gives

$$-\phi_i V = I_i R_i + \frac{Q_i}{Ci} \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$= \frac{dQ_i}{dt} R_i + \frac{Q_i}{C_i} \quad . \quad . \quad . \quad . \quad (12a)$$

where ϕ_i is the fractional rotation of the potentiometer associated with C_i;

therefore

$$\frac{dQ_i}{dt} = -\frac{Q_i}{R_i C_i} + \frac{\phi_i V}{R_i} \quad . \quad . \quad . \quad . \quad (13)$$

Substituting $I_i = dQ_i/dt$ in eqn. (11),

$$C_0\frac{dV}{dt} = \frac{\phi V}{R_0} + \sum_{i=1}^{i=12}\left(-\frac{Q_i}{R_i C_i} + \frac{\phi_i V}{R_i}\right) \quad . \quad . \quad (14)$$

or

$$C_0\frac{dV}{dt} = \frac{\phi V}{R_0} - \sum_{i=1}^{i=12}\frac{Q_i}{R_i C_i} + \sum_{i=1}^{i=12}\frac{\phi_i V}{R_i} \quad . \quad . \quad . \quad (15)$$

Comparison of eqns. (6) and (15), and of eqns. (7) and (13) shows that the reactor behaviour is correctly represented with the following identities:

$$V = P$$
$$C_0 = l^*$$
$$\phi/R_0 = \delta$$
$$\phi_i/R_i = \beta_i$$

$R_i C_i = 1/\lambda_i$ = Mean lifetime of ith delayed-neutron group

$$Q_i = X_i$$
$$\sum_{i=1}^{i=12}\frac{Q_i}{R_i} = \sum_{i=1}^{i=12}\beta_i = \beta$$

[The discussion on the above paper will be found on page 447.]

621.317.79 : 681.142 : 539.712.4

Paper No. 2338 M
Mar. 1957

THE APPLICATION OF ANALOGUE METHODS TO COMPUTE AND PREDICT XENON POISONING IN A HIGH-FLUX NUCLEAR REACTOR

By G. J. R. MacLUSKY, B.Sc.

(*The paper was first received* 10*th July*, 1956, *and in revised form* 7*th January*, 1957. *It was published in March*, 1957, *and was read before a joint meeting of the* MEASUREMENT AND CONTROL SECTION *and the* SUPPLY SECTION 12*th March*, 1957, *held in conjunction with the* BRITISH NUCLEAR ENERGY CONFERENCE.)

SUMMARY

The process of xenon poisoning, and its relation to the operation of a high-flux nuclear reactor, are discussed. An analogue computer which gives the present value of xenon poisoning, and an associated predictor which can be used to plan future reactor operations, are described. This computer-predictor is to be used as a reactor operator's tool.

LIST OF PRINCIPAL SYMBOLS

c_I' = Concentration of iodine 135, atoms/cm^3.
λ_I = Decay constant of iodine 135.
ϕ = Thermal-neutron flux.
γ_I = Fractional yield of iodine 135 as a direct fission product.
Σ_f = Macroscopic cross-section of the fuel in the reactor.
c_X' = Concentration of xenon 135, atoms/cm^3.
c_{X0}' = Steady-state concentration of xenon 135, atoms/cm^3.
λ_X = Decay constant of xenon 135.
σ_X = Microscopic thermal-neutron absorption cross-section of xenon 135.
γ_X = Fractional yield of xenon 135 as a direct fission product.
P = Reactor power.
c_X = Loss of reactivity due to xenon 135.
c_I = Concentration of iodine 135 relative to xenon 135.
V_I = Computer output voltage proportional to c_I.
V_X = Computer output voltage proportional to c_X.

(1) INTRODUCTION

The reactivity of a nuclear reactor may be affected by the growth within it of fission products which have large neutron-capture cross-sections. The most important of these fission products is xenon 135, and previous papers by Moore[1] and by Cox and Walker[2] have discussed the importance of xenon poisoning in high-flux reactors, and the way in which, when such a reactor is shut down, the loss of reactivity will slowly increase to many times the equilibrium value before it finally decays. For example, in a heavy-water-moderated enriched-fuel reactor with a peak neutron flux of 10^{14}n/cm^2/sec the equilibrium value of xenon poisoning has been calculated as 5·2% in reactivity. This will increase during shut-down at an initial rate of 2·9% per hour, reaching a peak of 18% after 11 hours and then requiring a further 28 hours to fall through its original equilibrium value. Such variations in reactivity are very large compared with the reactivity adjustments which are directly under control, such as those effected by control-rod position, effective reactor volume or location of materials being irradiated. Consequently, the changes which occur in xenon poisoning will at all times affect the control of such a high-flux reactor, and may under some circumstances completely outweigh the effect of all other controls.

(1.1) Effect of Xenon Poisoning on Reactor Operations

Suppose that a reactor has been operating at some steady conditions of power and gross reactivity (i.e. reactivity as set by control rods, etc.) and is then shut down for an appreciable period. Before it can be started up again, the gross reactivity must be increased beyond the original operating conditions by an amount corresponding to the increase in xenon poisoning which has taken place during the interval. For rapid starting-up the operator may need information regarding the gross reactivity at which the reactor may be expected to become critical again. But in any reactor there are definite limitations to the rate of increase of gross reactivity and on the excess reactivity available, and the rate of build-up of zenon poisoning therefore imposes a limit on the duration of any controlled shut-down. Any longer shut-down will allow the poisoning to increase so far that starting-up is no longer possible (colloquially, the reactor has 'poisoned out') and it is necessary to wait for the eventual decay of the xenon to a suitable level, which may be some 40 hours later.

Similar considerations will apply in the event of a rapid reduction in reactor operating power, since after the initial small decrease in gross reactivity (to produce the required rate of fall of power) it may be necessary to provide for an appreciable rate of increase of gross reactivity merely to maintain the new steady power level because of the rapid build-up of xenon poisoning. For a rapid increase of power the converse will apply, and there may even be the possibility of an uncontrolled surge in power. These considerations will become especially serious in highly-rated mobile power reactors, since their application may enforce rapid changes in output power, and some form of computing equipment would appear to be an essential tool for use by the reactor operators.

(1.2) Requirement for a Computer

If it is assumed that the reactor has operated at one assigned power for sufficiently long for the xenon poisoning to reach equilibrium, the subsequent course of poisoning during shut-down can be calculated and made available to the operators in graphical form.

But the complexity of the data required increases rapidly if it is required to allow for many different operating powers, and also an experimental reactor may well be operated in conditions more variable than would apply to a production reactor, so that the poisoning may not have time to reach any equilibrium value.

For these reasons, an electronic analogue computer has been developed to be used in the instrumentation of high-flux reactors. It is to be supplied continuously with the value of the reactor power and will compute from it the amount of xenon poisoning.

(1.3) Requirement for a Predictor

There may also be circumstances in which the reactor operator must choose between alternative courses, e.g. between shutting down the reactor rapidly, attempting to clear some fault in a limited time and then restarting before the reactor has 'poisoned out', or continuing to run the reactor at some reduced power level until the poisoning has decreased, and then shutting down for a longer time. Clearly such decisions must involve

Mr. MacLusky is at the U.K.A.E.A. Atomic Energy Research Establishment.

an estimate of the future course of poisoning in various circumstances, and so the equipment which has been built includes a predictor section. This takes its initial conditions from the associated computer, and the assumed future reactor power is set in manually. The predictor, operating on a time-scale accelerated by a factor of 60 or 3 600, predicts the poisoning over some future period.

(2) PHYSICAL BASIS*

The relevant radioactive processes are

$$5{\cdot}6\%\text{ yield } Te^{135} \xrightarrow[1\text{ min}]{\beta} I^{135} \xrightarrow[6{\cdot}7\text{ h}]{\beta} Xe^{135} \xrightarrow[9{\cdot}2\text{ h}]{\beta} Cs^{135} \xrightarrow[2{\cdot}1\times10^{6}\text{ yr}]{\beta} Ba^{135}$$

$$0{\cdot}3\%\text{ yield} \longrightarrow \text{(neutron capture)} \longrightarrow Xe^{136}$$

Since the half-life of tellurium 135 is much shorter than that of iodine 135, there is negligible error in treating the latter as the direct fission product, and the equation for the concentration of iodine becomes

$$\frac{dc_I'}{dt} = \gamma_I \Sigma_f \phi - \lambda_I c_I' \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The corresponding equation for the xenon concentration is

$$\frac{dc_X'}{dt} = (\gamma_X \Sigma_f - \sigma_X c_X')\phi + \lambda_I c_I' - \lambda_X c_X' \quad . \quad . \quad (2)$$

Equilibrium poisoning corresponds to the steady state when $dc_X'/dt = 0$ and $dc_I'/dt = 0$,

and therefore $$c_{X0}' = \frac{\Sigma_f(\gamma_X + \gamma_I)\phi_0}{\lambda_X + \sigma_X\phi_0} \quad . \quad . \quad . \quad . \quad (3)$$

At low values of ϕ_0 the left-hand term in the denominator will predominate and the xenon concentration will increase linearly with flux. For $\phi_0 > 10^{14}$, the right-hand term will predominate and there will be no further increase of xenon concentration.

These equations refer to the concentrations of iodine and xenon at any one point in the reactor. It can be shown by perturbation theory that loss of reactivity due to xenon is given by

$$c_X = \frac{\int c_X'\phi^2 dV}{\int \phi^2 dV} \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where the integrations are carried out over the whole volume of the reactor.

If eqns. (1) and (2) are multiplied by ϕ^2 and integrated over the whole reactor, they can be reduced to

$$dc_I/dt = Q_1 P - \lambda_I c_I \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$dc_X/dt = Q_2 P - Q_3 P c_X + \lambda_I c_I - \lambda_X c_X \quad . \quad . \quad (6)$$

where c_i and c_x are averaged over whole reactor, and Q_1, Q_2, Q_3 are constants dependent on the nature of the reactor fuel and the neutron flux averaged over the whole reactor.

However, this simplification depends on the use of an approximation in integrating the $\sigma_X c_X'\phi$ term in eqn. (2), and this approximation ceases to be valid if there is a large variation in the flux over the region concerned. For the reactors under immediate consideration the flux variation across the core is relatively small, and so it is possible to apply eqns. (5) and (6) to the whole reactor, resulting in a simple analogue. For reactors having very non-uniform flux distributions it is necessary to consider the reactor as being divided into a number of separate regions, chosen so that

(*a*) The flux variation across each region is relatively small.
(*b*) The integral of $\phi^2\,dV$ across each region is the same.
(*c*) The boundaries of the regions are contours of constant flux.

Eqns. (5) and (6) are then solved separately for each region with different values of Q_1, Q_2 and Q_3, and the reactivity effect for the whole reactor is proportional to the sum of the resulting solutions for c_x.

(2.1) Analogue System

A block schematic of the analogue system used is given in Fig. 1, the upper half being the computer section and the lower

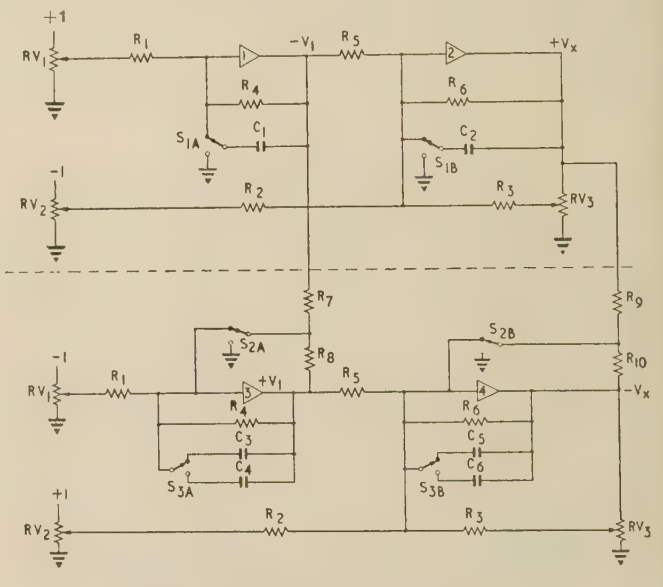


Fig. 1.—Schematic of analogue system.

the predictor section: components having the same function are given the same number in both sections. In both sections, RV_1, RV_2 and RV_3 are ganged controls, whose fractional rotation can be denoted by θ. Then, for the computer section, with S_1 in the position shown.

$$\frac{dV_I}{dt} = \frac{\theta}{R_1C_1} - \frac{V_I}{R_4C_1} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

making the usual assumption that the input of the amplifier is at earth potential.

Similarly,

$$\frac{dV_X}{dt} = \frac{\theta}{R_2C_2} - \frac{\theta V_X}{R_3C_2} - \frac{V_X}{R_6C_2} + \frac{V_I}{R_5C_2} \quad . \quad . \quad (8)$$

Comparing eqns. (7) and (8) with eqns. (5) and (6), it is seen that the analogue correctly represents the required systems with the following identities:

$$\lambda_I = 1/R_4C_1 = 1/R_5C_2$$
$$\lambda_X = 1/R_6C_2$$
$$P = \theta$$
$$Q_1 = 1/R_1C_1$$
$$Q_2 = 1/R_2C_2$$
$$Q_3 = 1/R_3C_2$$

When the switch S_1 is in the position which connects C_1 and C_2 to earth, the output voltages will quickly reach values corresponding to the steady state amounts of iodine and xenon, for

* This Section has been in part adapted from an unpublished communication by P. R. Tunnicliffe; a similar treatment is given in Reference 4.

the particular value of P. This is used in initial adjustment of the computer. The ganged potentiometer RV_1, RV_2, RV_3 is, in the computer section, driven directly by a recorder which follows the reactor power at all times.

In the predictor section, R_7 and R_8, R_9 and R_{10} are matched resistor pairs, and the time-constants R_8C_3 and $R_{10}C_5$ are made very small. So, with S_2 in the position shown (standby), the output voltages of amplifiers 3 and 4 will match those of amplifiers 1 and 2, apart from a reversal of sign.

By making $C_3 = 60\ C_4$ and $C_5 = 60\ C_6$, and with $60\lambda_i = 1/R_4C_3$ and $60\lambda_X = 1/R_6C_5$ the two positions of switch S_3 will cause the predictor section to operate in a time scale accelerated by factors of 60 and 3 600 respectively. The ganged controls RV_1, RV_2 and RV_3 are set manually in the predictor section according to the estimated future power, and when S_2 is changed to the 'predict' position this section will start its computation from initial values determined by the outputs of amplifiers 1 and 2.

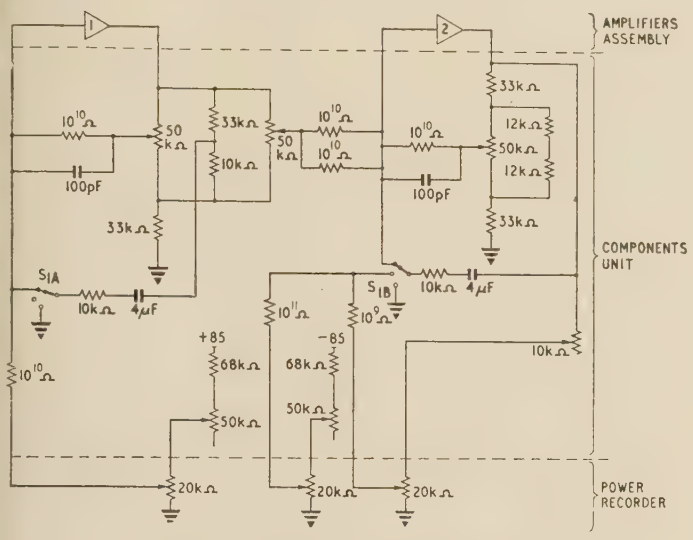


Fig. 2.—Circuit of computer section.

(2.2) Details of Computer and Predictor Circuits

The four amplifiers used are of the type described in the companion paper (see p. 433), except that the vibrator stabilizing systems are disconnected in amplifiers 1 and 2. In this application the impedance level of the circuits is so high ($\simeq 10^{10}$ ohms) that the additional spurious input current produced by the vibrator system is more serious than the input drift in the absence of vibrator stabilizing. The characteristics of the amplifiers are summarized below.

	With vibrator stabilizing	*D.C. amplification alone*
Gain	4×10^6 at d.c. 4×10^3 from 1 to 100 c/s.	4×10^3 from d.c. to 100 c/s.
Drift	$\pm$ 50 μV	$\pm$ 5 mV, short term $\pm$ 50 mV, long term
Current at input	$< 5 \times 10^{-11}$ amp	$< 5 \times 10^{-12}$ amp.

Figs. 2 and 3 show the circuits used for the computer and predictor sections, which are arranged to employ standard components and to permit adjustment to cover component tolerances and drifts. The high-value resistors employed are of the glass-sealed type, and since they have a resistivity coefficient of 0·25% per volt, averaged over 0–100 volts, the voltage applied across them is limited to 15 volts. The capacitors used

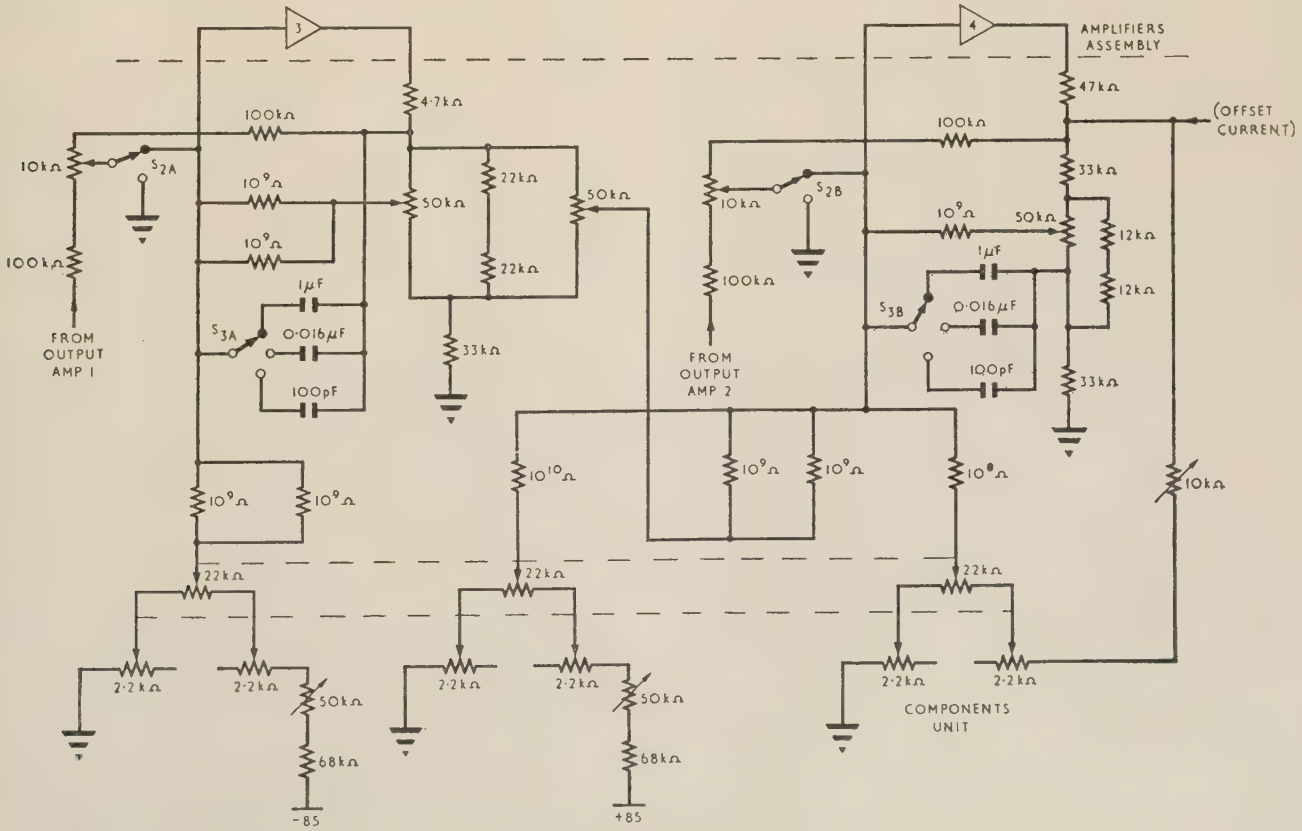


Fig. 3.—Circuit of predictor section.

have polystyrene-film insulation, and have a time-constant (in the absence of shunting external resistance) of about 500 hours. Switch S_1 has an intermediate position in which the capacitance is completely isolated, and this is used when it is required to recalibrate the equipment without appreciable loss of stored information. The 10-kilohm resistor in series with the $4\,\mu$F capacitor is necessary to prevent the oscillations which would arise from capacitance loading of the amplifier output in the 'earthed' position of S_1.

In the predictor section, the $1\,\mu$F and $0{\cdot}016\,\mu$F capacitors are adjusted by trimming capacitors to give a capacitance ratio of exactly 60 : 1, and a third position of S_2 provides a very short time-constant used during calibration. RV_{19} and RV_{16} are switched chains of accurate wire-wound resistors, giving respectively 10% and 1% steps in predicted reactor power.

In normal operation a self-balancing potentiometer-type recorder is driven from the output of the computer section, and switching is provided for full-scale ranges of 5, 10 and 25% in reactivity, corresponding to 4, 8 and 20 volts at the output of amplifier 2. The predictor output is displayed on a moving-coil recorder whose chart travels at 0·5 in/sec while S_2 is in the predict position. This recorder is normally connected to the output of amplifier 4, with series resistors switched to give ranges of 5, 10 and 25% in reactivity. Provision is made for electrically shifting the zero of the predictor output by up to 20% in reactivity. Since the impedance of the recorder is much lower than that of the computing feedback resistors, an offset current is supplied from a high-impedance source, as shown, to bring the computing networks to the required voltage while leaving the amplifier output at earth potential.

(2.3) Adjustment and Calibration

Previous observations on the type of glass-sealed high-value resistors employed had shown very slow drifts of up to 10% in resistance, and further random variations of $\pm 1\%$ with periods of about a few weeks. Provision was therefore made for rapid recalibration of the instrument at, say, daily intervals without disturbing the charges stored on C_1 and C_2, with S_1 in its central ('test') position, which isolates these capacitors. This requires that the input and feedback resistors in both computer and predictor sections shall be connected to various test voltages in a predetermined routine, and this is carried out by a 14-pole 7-way switch not shown in the diagrams. The two recorders are connected to either iodine or xenon output points as required, and at each stage the computer and predictor sections are recalibrated together.

To simplify the routine, only the ratios of R_1/R_4, and R_2, R_3, R_6/R_5 are reset in this way. In the initial calibration the ratios R_7/R_8 and R_9/R_{10} are set, and the time-constants R_4C_1 and R_6C_2 are made to give the correct values of λ_I and λ_X respectively.

(2.4) Mechanical Arrangement

As indicated in Figs. 2 and 3, the major computing components and controls for both computer and predictor sections are located in a single components unit. The four amplifiers are mounted as a separate assembly, and a third unit contains power supplies. Since the high-value resistors may have a temperature coefficient of 0·43% per deg C, they are mounted in a temperature-controlled section within the components unit.

The only controls which are normally accessible are the 'standby predict' switch S_2; the two decade switches which set predicted power, RV_{16} and RV_{17}; and S_3, which determines the time scale of prediction. The remaining preset controls are in two groups, used in a routine recalibration or in initial adjustment respectively. As a safeguard against accidental destruction

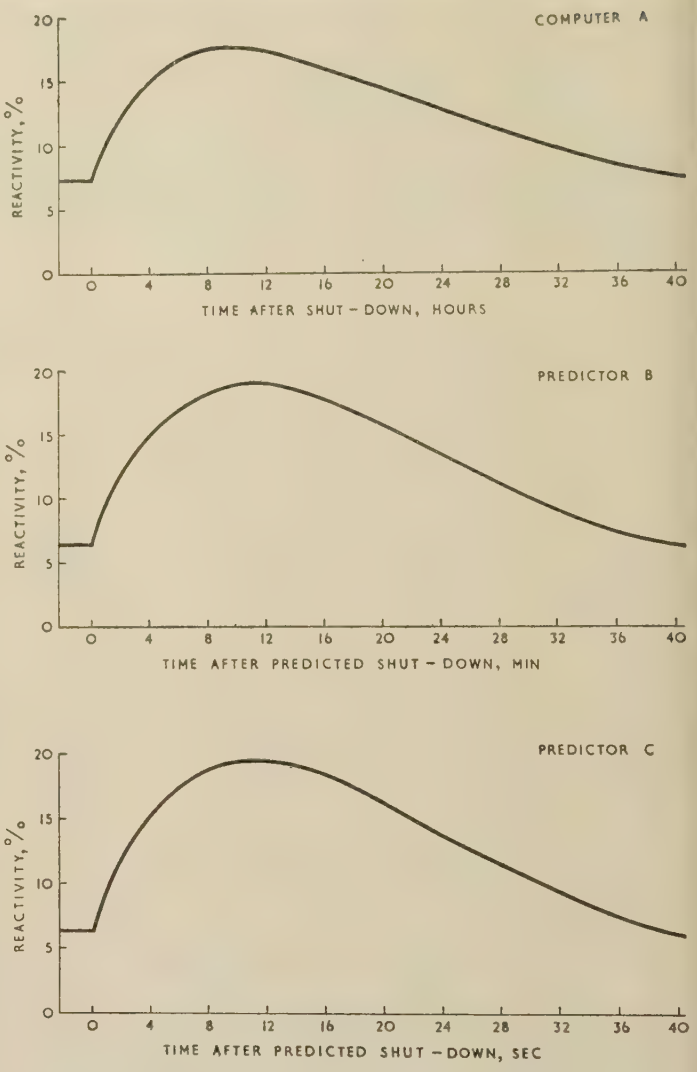


Fig. 4.—Response of computer and predictor.

of stored information, a mechanical interlock allows S_1 to be put into the 'earthed' position only when the second group of controls are in use.

(3) OPERATION

The computer section has been operated in conjunction with an artificial input representing the variations in reactor power. Fig. 4 shows the build-up and decay of xenon poisoning following a complete shut-down of the reactor. Figs. 4(*b*) and 4(*c*) are the corresponding predictor outputs on the 'seconds to minutes' and 'seconds to hours' time-scales respectively.

(4) REFERENCES

(1) Moore, R. V.: 'The Control of a Thermal Neutron Reactor', *Proceedings I.E.E.*, Paper No. 1431M, December, 1952 (**100**, Part 1, p. 90).

(2) Cox, R. J., and Walker, J.: 'The Control of Nuclear Reactors', *ibid.*, Paper No. 2068M, March, 1956 (**103**, Part B, p. 577).

(3) Glasstone, S., and Edlund, M. C.: 'Elements of Nuclear Reactor Theory' (McGraw-Hill, New York, 1955).

(4) Schultz, M. A.: 'The Control of Nuclear Reactors and Power Plants' (McGraw-Hill, New York, 1955).

DISCUSSION BEFORE A JOINT MEETING OF THE MEASUREMENT AND CONTROL SECTION AND THE SUPPLY SECTION, IN CONJUNCTION WITH THE BRITISH NUCLEAR ENERGY CONFERENCE, 12TH MARCH, 1957

Mr. R. W. Sutton: In considering the dynamic behaviour of a complex system such as a nuclear power plant, it is necessary first to establish some form of mathematical model to represent the system. This will usually take the form of a number of first-order differential equations, some or all of which may be non-linear, and in order to obtain solutions quickly there is no doubt that some form of computer is required. The author puts forward two arguments in favour of the analogue type of machine, the second of which, namely the facility for changing the system parameters and for direct observation of results, I endorse. I do not think, however, that lack of accurate knowledge of the system parameters justifies a lower accuracy in computing answers. One should attempt to maintain the highest possible accuracy of computation at all times, and computers should be designed with this in mind.

Another important point in connection with system design is the model chosen to represent the system. Obviously one wants the best possible model, because it reduces the discrepancy between the computed answers and those from the plant when it is manufactured. However, generally speaking, such a model will be very complex and, because of non-linearities, e.g. the variation of heat-transfer coefficients with coolant velocity, may be somewhat difficult to realize in an analogue machine; but this is not sufficient justification for rejecting the analogue machine and using a digital computer for studies involving a wide range of system parameters.

The approach which should be made is to subdivide the system, as the author, in fact, does to some extent with his kinetics, control and thermodynamics sections; but one should go further and study individual components, such as the heat exchanger, which is represented reasonably well by three partial differential equations. One can solve these by removing the space variable by finite-difference methods, and then examine to find the simplest representation which adequately describes the behaviour of the heat exchanger. Similarly, by examining the reactor behaviour under various conditions of changing inlet temperature, and reactivity changes due to control-rod motion, the important features of the behaviour may be retained with the simplest model. By applying this technique to the whole system it is possible to reduce the complexity of the analogue, and in doing this one will also learn a great deal about the behaviour of the individual components; I think this is very important.

These general remarks lead to a number of particular points, the first of which deals with the condition arising when one must take into account both temporal and spatial variations. The author suggests division of the system into a number of zones and averaging over each zone; has he any particular reason for doing this, as opposed to using standard finite-difference methods?

On the question of simplifying the problem before examining the complete system, I wonder what governed the choice of 40 amplifiers for this particular computer. I think that 30 would have given fairly adequate representation of the reactor system if one carried out the simplification suggested. Again, if there are 40 amplifiers, more than 40 coefficient units are needed, since many of the amplifiers would have more than one input.

The papers mention that a time-constant range of $10^4 : 1$ is available. I am not sure that it is really essential, because a time-constant of 0·01 sec anywhere in the system can be neglected in comparison with those of 100 sec.

Can the author gives details of the comparative accuracies of servo potentiometers and electronic multipliers, particularly for the type of problem which he mentions on overall plant studies, where power will be changing relatively slowly?

With regard to the size of the computer, if a further reactor-kinetics section were added, it could be arranged for two problems to be tackled simultaneously and the efficiency of usage would be increased.

In paper No. 2338 M, under what conditions is eqn. (4) valid? In eqn. (6), Q_3 is taken to be a constant; to what extent is this justified, and is it proposed to be able to change this? Finally, have the effects of temperature on the production of poisons been considered?

Mr. I. Wilson: An analogue computer built at Harwell in 1953 would compute neutron-flux transients covering the wide effective range experienced during the starting-up of most reactors by using a 'clamping' technique similar to that described by the author. When techniques enabling analogue computation to be applied to reactor-kinetic studies had been developed, a larger machine was built in 1954 with a view to handling the kinetics of complete power plants. A reactor-kinetics section computes neutron flux, log flux and flux 'doubling time' over a continuous span of five decades (100 dB), and 40 operational amplifiers and several servo multipliers are available. For some problems it was found necessary to develop an accurate fast electronic multiplier using the time-division principle. Zero errors are of the order of 0·02% of full output (100 volts) and phase lag is about 1° at 100 c/s.

Transport delays are simulated by a charge-storage system employing 100 plastic-film capacitors mounted on a rotating drum (Fig. A). Each capacitor receives a charge proportional to the instantaneous value of the input signal, the charge being measured by a reading-out circuit after an interval approximately equal to the time of one drum revolution. The velocity servo controlling the drum speed enables the delay time to be varied independently between one and several hundred seconds.

To achieve realism in reactor simulation for training purposes, it is necessary to obtain as wide a dynamic range as possible in the neutron-flux computing circuits if the all-important starting-up procedure is to be demonstrated. The display of log neutron flux and doubling time over this range is essential, and current techniques enable these quantities to be computed continuously over a range of six decades (120 dB). A wide-range logarithmic d.c. voltmeter* was developed for these applications, and its principle of operation is explained in Fig. B.

Training simulators are in operation at Harwell and at Calder Hall. The Calder operations training simulator is a large machine which simulates all important aspects of plant operation for a Calder-type station.

Mr. N. Rydell (*Sweden*): Paper No. 2338 M mentions that a vibrator input cannot be used with the unit amplifier, owing to the high impedance levels necessary in the associated network. However, the standard types of d.c. amplifier without drift correction tend to have a fairly large drift, so we have developed a unit amplifier (Fig. C) with a gain of 2×10^5 from d.c. to 20 c/s, a drift of less than 1 mV/h and a current at the input of less than 10^{-13} amp. The quoted drift figure is, in fact, somewhat conservative, but we hope to reduce it further by improved technique. The most important feature of this amplifier, which has no automatic drift correction, is that the first two stages consist of directly-heated valves, with a 10 mA filament supply obtained from the well-regulated anode line. This method of

* WILSON, I., and COX, R. J.: 'Improvements in and Relating to Voltmeters', British Patent Specification No. 27468.

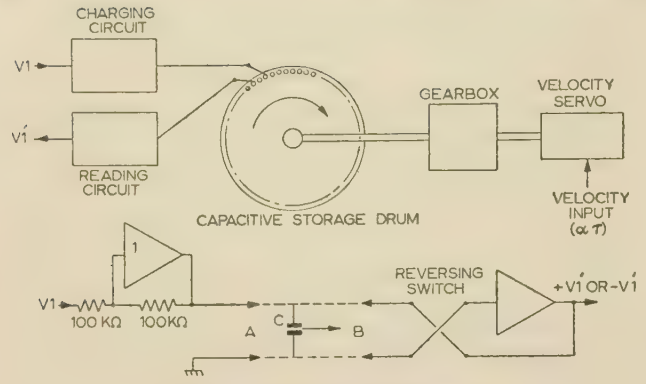


Fig. A.—Delay-line analogue.

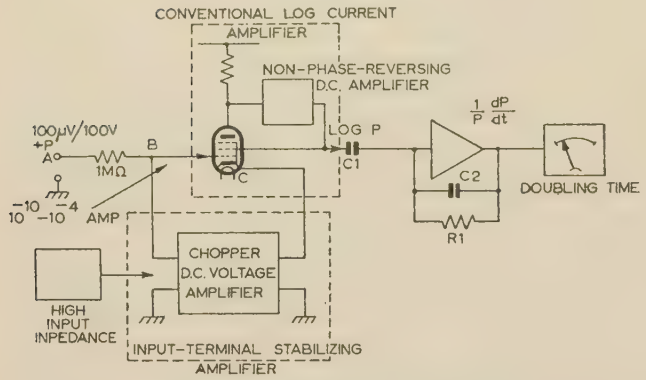


Fig. B.—Linear-logarithmic convertor and period meter.

cathode heating gives the valve an extremely low drift, although to preserve it in an amplifier the filaments must be heated before anode potential is applied and the grid must be prevented from going positive during any time of its operation—which makes it necessary to build into the analogue computer a special switching-on programme. However, with these precautions the amplifier is very reliable and well suited for analogue computers, particularly where long time-constants appear.

Servo-multiplication is the best method of reactivity simulation and it is better to connect P than δK across the potentiometer; but the elaborate methods employed to correct for the slowness in the servo are unjustified. The reactivity input has two components—the internal and external reactivity changes, i.e. those coming from temperature changes in the core materials and those simulating the effects of control rods, etc. The internal changes are always slow enough to be closely followed by the servo; the external changes are often slow under normal operating conditions, but may be fast and large, e.g. at shut-down. For engineering studies of reactor performance a simple system without compensations is fast enough, but for more fundamental studies of transfer functions, etc., where accuracy is really important, one should seek much greater accuracy in the simulation of external reactivity changes by shifting resistor values from output to input of the power amplifiers. If a corrective network is complicated and still does not give all the necessary correction it should be avoided, because it is important that the many sub-units in a big computer should function in a simple way which is easy to set and check.

With the same method of multiplication we have found that the resolution of a potentiometer of 0·1% linear accuracy is too poor when full-scale reactivity changes are of the order of 5–15%. This poor resolution appears as bumps in the power-output characteristic. Has the author encountered this difficulty in his simulator?

The values of the physical parameters in the poison equations (Paper No. 2338 M) seem to be old, and new measurements indicate a yield of iodine of more than 6%. What is the accuracy of the poison prediction, and what is the reactor operator's experience with this system as a tool?

Mr. R. J. Smith: What range of transit times does the author obtain with his magnetic-tape transit-time delay unit? I am proposing to use a unit similar to that described by Mr. Wilson and expect to achieve a range of delay times of 2–60 sec.

I like the author's arrangement for the reactor-kinetics section, particularly the simple delayed-neutron group arrangements. When this circuit is used over a wide range of power, involving the clamping technique described, is there not some risk of the relay contacts changing over at slightly different times, thus disturbing the capacitor voltages?

The arrangement for introducing frequency compensation on the reactivity potentiometer is elegant, and one might convert it from the essentially linear system which the author first describes to a non-linear one in which the product of the instantaneous power and the reactivity is obtained by servo-driving the potentiometer RV_{21} so that its angle of rotation corresponds to the reactor power. When studying automatic power-control systems over a limited range of power, I have found the resolution of the reactivity potentiometer troublesome, but have overcome this by servo-driving the potentiometer so that the reactor power controlled the angle of rotation and the reactivity voltage was supplied across the potentiometer. With this arrangement the frequency response of the servo system is not so important, because the power signal which controls the movement is not changing as rapidly as the reactivity.

As a user of analogue computers I am interested in any proposal which limits the number of amplifiers and other components; but although we have 80 amplifiers in our computer, together with associated coefficient units, we have found them all to be necessary. I agree that it is desirable to simplify the analogue representation of the various sections of a nuclear power station by commencing with a highly accurate and extensive representation, possibly taking spatial variations into account by finite-difference methods, and then introducing successive approximations up to the point at which errors become significant. In following this technique on the heat exchange occurring in an extended fuel-element channel we have found it necessary to use up to 80 amplifiers in the initial stages of the investigation.

In Paper No. 2338 M it might be of value to arrange the combined computer-predictor so that it is possible to feed into it a programme of power changes rather than merely one power change as the author describes. There would seem to be no fundamental difficulty about this, but has the author considered it?

Dr. J. Weill (*France*): In comparison with the Harwell analogue computer I should like to describe the electronic division of the French Atomic Energy Commission analogue laboratory at Saclay. This centre comprises four machines, with 90 d.c. amplifiers, 144 potentiometers, 15 multipliers, 4 function generators, one delayed-function unit, one 2-dimensional space-potential simulator, 16 recorders, and one time calculator with automatic on and off. Fig. D shows the complete arrangement of this installation.

However, the important thing about analogue computers is what they can do, and what calculations can be performed with their aid. At Saclay we have done the following calculations:

Start-up and shut-down of reactors.
Influence of the speed of control rods and safety rods.

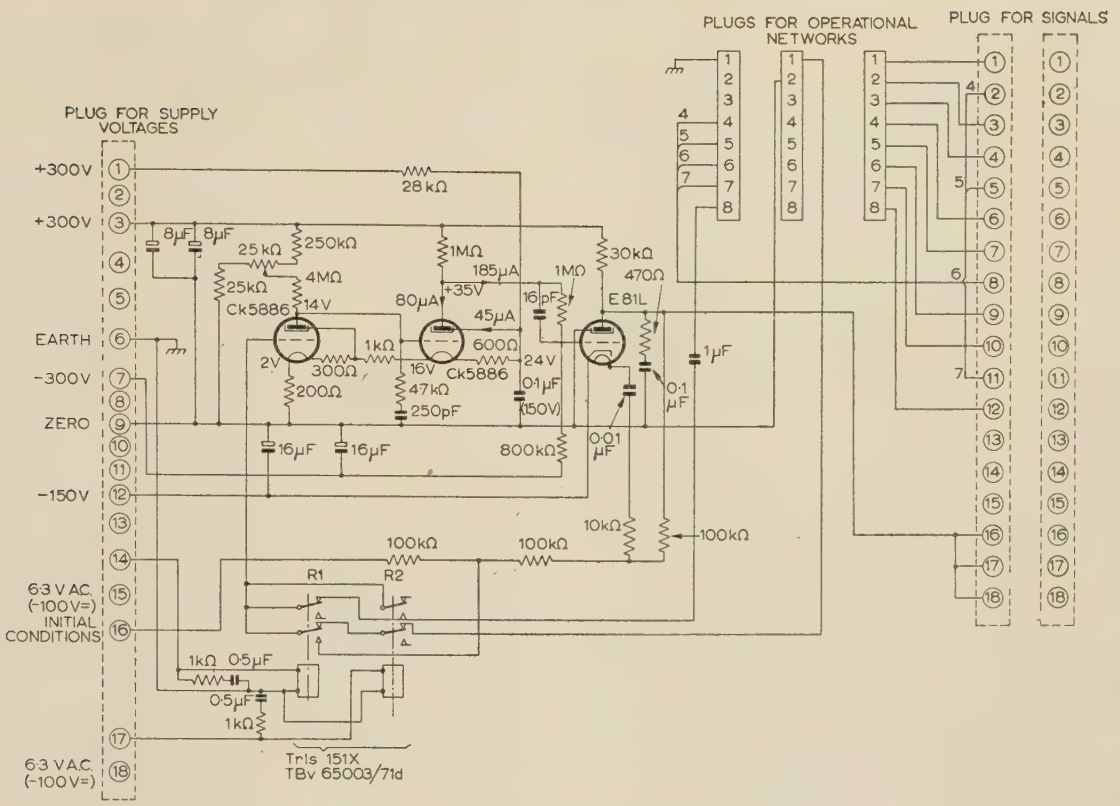


Fig. C.—An operational amplifier.

Troubles in cooling around fuel rods cooled by heavy water.
Complete simulation of a reactor and its thermal circuit and utilization.
The effect of voids in an homogeneous plutonium reactor.

In a completely different field of application we used the computer to study the phase-injection effects in a linear accelerator.

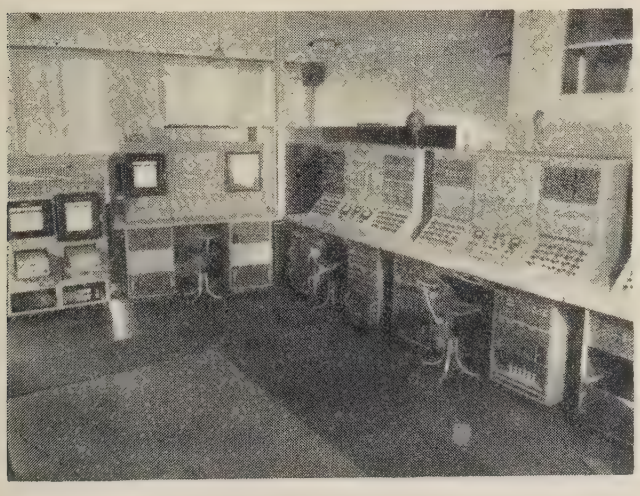


Fig. D.—The analogue-computer laboratory at Saclay.

Monsieur C. Caillet (*France*): Because of their inaccuracies, d.c. amplifiers cannot give direct solutions when the numerical values involved vary over a large range, and thus it becomes necessary to break down the problem into a series of elementary calculations. Such step-by-step methods are inconveniently long, especially when the simulated equations are not homogeneous; any change of scale modifies the coefficients and so necessitates resetting the potentiometers for each step. The problems arise particularly in the study of 'radiolyse voids' and temperature effects in a homogeneous reactor when a step function of positive reactivity is injected from the steady state, for the ratio between the maximum and actual powers may reach 10^6–10^9. A preliminary mathematical method* has recently been evolved to eliminate the step-by-step process in the study of the kinetics of the homogeneous reactor at Saclay.

Monsieur P. Braffort (*France*): The non-linearity of the equations in the prediction of xenon poisoning in a nuclear reactor necessitates an analogue computer, and there are two possibilities available, namely

(*a*) A real-time simulator, with the instantaneous value of the xenon concentration fed correlatively to the continuous signal of reactor power.†

(*b*) A simulator with a memory, with the power recorded in real time and available over a short period; instead of computing continuously as the nuclear power varies, the apparatus determines the xenon concentration only when required, using the nuclear-power figures stored in the memory.‡

Method (*b*), which is used at Saclay, eliminates the need for capacitors exceeding 10 μF, whose leakage resistance approaches that of the resistors in the analogue circuit. Moreover, the computer may be used for prediction: if a hypothetical nuclear-power evolution is introduced, the simulator will indicate the trend of the xenon kinetics for the ensuing hours. For different laws of motion, different possibilities of the xenon trend will arise, and there is some possibility that the apparatus can be

* Caillet, C., and Chazal, G.: 'Effet de radiolyse et de température dans un réacteur homogène', Rapport Intérieur CPLA 71.
† Goulding, F. S., and Ward, A. G.: C.R.E.L., 597, 1955.
‡ Braffort, P.: 'Ensemble prédicteur programmeur pour l'effet xenon dans un réacteur nucléaire', Étude MCLA 31.

modified to select an optimum programme for itself;* thus it will become part of the programmer.

Messrs. P. Braffort and **C. Caillet:** We have analysed the different steps of the resolution of the system of equations by which the dynamics of a circulating-moderator nuclear reactor, with recovery of energy, is represented, and find that the physical data relating the setting of different organs of the whole system give the possibility of building the corresponding 'physical diagram'.

In order to establish the analogue diagram, it is convenient to start from the mathematical equations and use them according to rules that have now become classical, and we would insist upon this interest in the 'mathematical diagram' prior to the construction of the analogue diagram, since the production of a partially ordered set structure allows the formulation errors to be found easily.

The difference between the physical diagram and the mathematical diagram is rather similar to that between the block diagram and the signal-flow diagram in the theory of automatic feedback control systems; moreover, there are numerous common points between the design of the mathematical diagram and that of the signal-flow diagram.

The resolution of problems studied by analogue computers sometimes shows unsteady states, which are particularly dangerous when nuclear reactors are being considered. Here again, the use of systematic reduction techniques very similar to those used in the theory of signal-flow diagrams allows one to reduce the number of possible diagrams to a few typical cases and quickly to discover the steady states for the reactors.

Mr. R. Potter: In the computer described there are quite a number of direct external connections between amplifiers and coefficient boxes. I believe that the system of centralized interconnections by means of interchangeable boards—which is the system being adopted on some computers—offers a number of advantages, including easier checking of circuits, and it is also possible for engineers and mathematicians to use the computer directly, which is valuable.

I agree that insufficient attention has been given to the problem of spatial variations; although a great deal of work has been done, notably by the late Gerhardt Liebmann, there is a definite field for research in the combination of passive and active networks to solve systems of partial and ordinary differential equations. The present system of making a problem fit the computer is not desirable and tends to become a habit. At Harwell we have recently succeeded in simulating heat flow by means of passive networks, and we have solved a set of partial and ordinary differential equations.

I wonder whether the checking procedure in setting up computers is as complete as it should be. The checking of steady states provides a partial check, but at least two different steady states need to be checked rather than one, which is the normal procedure. Some things, however, can be checked only by checking individual units and their connections. This is where the system of centralized interconnections and the simplicity of units is very desirable.

In Section 4.2 of Paper No. 2337 M the maximum reactivity available is given as $+1\%$ to -4%. For the modern reactors using enriched or pure uranium 235 it is certainly inadequate.

* BRAFFORT, P.: Communication au Congrès International de l'Automatique, Paris, 1956.

Mr. P. R. J. French: Fig. 9(*a*) of Paper No. 2337 M shows a condenser connected across the input resistor to an amplifier, and the author mentions the testing of this device—which is simulating the servo control of the reactor—by switching the demanded power, represented by a voltage applied to the input resistor. If the voltage is changed suddenly it appears that the condenser shunting the input resistor will be suddenly charged, and on the face of it one would expect the amplifier to be heavily saturated and to interfere with the correct recording of the subsequent transient.

The 10^{10}-ohm resistors mentioned in Paper No. 2338 M seem to be a nuisance in several ways: their resistance may change by 25% when 100 volts is applied across them, so that they would make a good device for non-linear bridges if they were not subject to drifts of up to 10%. Could not these resistors be eliminated by utilizing some of the amplification which is available in the d.c. amplifiers to increase the circuit time-constants electronically?

Mr. R. Postlethwaite: I am rather surprised that no speaker has mentioned work on an advanced time scale, for several have stated that they are using analogue computers for basic investigation and that quite often accuracy is not so important. It would therefore be more appropriate to obtain a full working knowledge of the system by using the advanced-time-scale technique.

To consider one aspect of the computer described, namely the amplifier, I am surprised to see that the author uses 33-megohm resistors in the first stage, since the lag at this point is a major limiting factor in frequency response. Surely it would be better policy to have electrometer valves in the input stage if a low grid current is required, and to use components which have been found to be reliable, i.e. components below 10 megohms and 1 megohm wherever possible?

In our application of analogue computing we deal with frequencies of up to 300 c/s in the unity time scale, and so the design of the amplifier is very different from that of the author. I appreciate that the author's time-constants are extremely long, but I am surprised to find 100-megohm resistors in temperature-controlled containers. This is not good practice, unless unity time scale has to be used, i.e. 'hardware' is included in the loop.

Mr. A. R. Owens: The possibility has been mentioned of using transistors in analogue computers, and thus obtaining the advantages of small size, increased reliability and lower power consumption. The main disadvantage of transistors is the rather serious dependence of certain parameters upon ambient temperature. In particular, the leakage current in a low-frequency junction transistor exhibits a drift of the order of 50μA for a change in temperature from 20 to 50° C. However, recent work has shown that these difficulties may be largely overcome, and the use of high-frequency transistors (which have lower leakage currents) as choppers, results in drift figures referred to the input terminal of 4×10^{-9} amp and 50μV for the above range of temperature.

Owing to the low input impedance ($\sim$500 ohms) of transistor amplifiers, it becomes convenient to regard the gain as a transfer impedance specified as output-voltage/input-current. A transfer impedance of 50 megohms is readily realizable, which allows a maximum feedback resistance of 0·5 megohm for 1% gain stability. A chopper-amplifier bandwidth of the order of 25 c/s has been achieved with a chopping rate of 2 kc/s. The above specification seems to indicate that the use of transistor amplifiers is feasible in some analogue-computer applications.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. G. J. R. MacLusky (*in reply*): I agree with Mr. Sutton that any computer should be designed to have the highest possible accuracy, provided that this is consistent with reasonable simplicity and reliability in view of the particular computing technique employed. My point was that the uncertainty of the final result will represent the addition of two sets of random quantities, namely the computing errors and the uncertainties in system parameters. Once the former have been made very small

compared with the latter, further effort to improve the computer accuracy is unrewarding.

The application of finite-difference methods will give a closer approximation to the true system behaviour than simple division into zones, but at the cost of appreciably greater complication in the analogue system.

A comparison of the comments by Messrs. Sutton and Smith and Dr. Weill shows that among computer users there are fairly wide differences of opinion as to the number of unit amplifiers required, and our choice of 40 was based on the best prediction we could make of future demand. There are spaces in the cabinets (and corresponding control-desk sockets) for the future addition of a further 20 amplifiers and coefficient units.

Operating experience has tended to support Mr. Sutton's suggestion that more coefficient units are required in proportion to the number of amplifiers, although it should be realized that 20 of the coefficient units are of the double-R type (Fig. 4), each of which can provide two amplifier inputs. Moreover, each amplifier has provision for a small plug-in unit which can contain two input and one feedback components, and these supplement the available coefficient units.

While agreeing with Mr. Potter that there would be many advantages in making all interconnections by means of detachable plug-boards, I feel that only the low-impedance points (amplifier outputs and coefficient unit inputs) should appear on such boards, in order to avoid spurious resistive leakages and capacitive couplings. This will cause some loss of versatility, since each amplifier input must have a fixed number of coefficient units permanently connected to it.

Regarding the suggestion that a wider range of reactivity variation is required, a reactivity of -4% gives a reactor power only 25 times the source term, and for any lower reactivity the reactor response can be calculated relatively easily. Conversely, a reactivity of $+1\%$ gives a period of only 3·5 sec in a large thermal reactor, where the prompt-neutron mean lifetime is about 1·4 millisec, and for a highly enriched fast reactor ($\tau \simeq 0{\cdot}1$ microsec) the corresponding period will be much less than 1/100 sec (see Fig. 3 in Reference 7). It should be emphasized that the limits above refer to the net reactivity applied to the reactor-kinetics section; its various inputs may separately cover far wider ranges, as decided when choosing the parameters in control and thermal sections.

I must dispute Mr. Sutton's suggestion that a 0·01 sec time-constant can always be neglected in comparison with others of 100 sec, since there may be circumstances where the system's initial response to a transient input is controlled by the former while the subsequent recovery is determined by the latter. Furthermore, in a feedback system having high gain, the presence of one short time-constant may modify the higher-frequency response, even though other large time-constants are present.

For slowly changing inputs the servo multiplier can easily achieve the same accuracy as the most precise electronic multiplier, and will have the advantage in terms of circuit complexity.

In reply to Mr. Rydell, the compensation for servo response in the reactivity section was intended to be used when simulating automatic power-level control and automatic start-up. To study the stability of such systems the loop gain and phase characteristics must be correctly simulated, particularly in the frequency range 0·1–10 c/s, even though the rates of change of reactivity are normally low. The accuracy of the servo compensation can be checked as described in the paper.

Mr. Smith has described an ingenious modification of the reactivity servo section which can be used to avoid troubles due to the resolution of the potentiometer. The magnetic-tape delay unit is intended to cover a range of delay times from 1 to 1 000 sec, with a time resolution of about 1 in 1 000. Still shorter delays could be obtained by reducing the distance between the magnetic heads, at the cost of a proportionate decrease in time resolution.

Disturbance of the charges stored on the capacitors in the reactor-kinetics section during clamping and reading operations is prevented by additional relays and interlocking contacts, not shown in Fig. 6, which ensure that the relays operate in a controlled sequence.

In reply to Mr. French I must explain that the system shown in Fig. 9(a) forms a portion of a feedback loop completed through the reactor-kinetics section. The voltage at the input is derived from that section and corresponds to the actual power, while the demanded power is determined by the position of the tap on R19/1; it is the latter which is changed suddenly. At all times the feedback current through the 10-megohm resistor balances the total input current through the RC combination, so that the amplifier does not overload, since its input is maintained near earth potential.

In reply to Mr. Postlethwaite, the 33-megohm anode load in the amplifier is used to control the frequency response of the amplifier, in conjunction with the 47 pF capacitor and 33-kilohm resistor between the anode and earth. Although the use of very-high-value resistors is undesirable in most equipments, it has been our experience that they can be used reliably if suitable precautions are taken to control their environment. In the xenon-poisoning computer, operation in real time was essential, and thus very long time-constants could not be avoided.

To deal with the other comments on the second paper, eqn. (4) is valid provided that one-group theory is applicable to the neutron distribution, as is the case for large thermal reactors. By treating the term Q_3 in eqn. (6) as a constant we are assuming that the reactor power (as read on the power recorder) is proportional to the average neutron flux, and that the capture cross-section of xenon is constant. Although the latter is known to vary with reactor temperature, this effect is not significant for the reactors to which this computer is to be applied. Q_3 is varied by the preset 10 kilohms (Fig. 2) and, if necessary, a further control operated by temperature changes could be included.

Using the amplifier gain to increase the effective time-constants, as suggested by Mr. French, would involve a large voltage step-down from amplifier outputs to the feedback resistors R4 and R6, and the extent to which the system was disturbed by amplifier drifts and changes in gain would be increased by the same factor.

621.384.611:573.534

Paper No. 2342 M
Mar. 1957

A CYCLOTRON FOR MEDICAL RESEARCH

By J. W. GALLOP, B.Sc.(Eng.), Member, D. D. VONBERG, B.Sc.(Eng.), Associate Member, R. J. POST, Associate Member, W. B. POWELL, M.A., Graduate, J. SHARP, B.Sc.(Eng.), Associate Member, and P. J. WATERTON, B.Sc.(Eng.), Associate Member.

(*The paper was first received 16th October*, 1956, *and in revised form 9th January*, 1957. *It was published in March*, 1957, *and was read before* THE INSTITUTION *21st March*, 1957.)

SUMMARY

The paper describes the design and construction of the Medical Research Council fixed-frequency cyclotron, which is to be used for research in radiotherapy and radiobiology and for the production of radioactive isotopes for diagnostic, therapeutic and medical research purposes.

The machine is designed to maintain an output of 200 μA of 15 MeV deuterons in an external beam under routine operating conditions.

Special care has been taken throughout the design to ensure ease of maintenance and reliable routine operation.

LIST OF SYMBOLS

m_0 = Rest mass of accelerated particle.
m = Relativistic mass of accelerated particle.
v = Velocity of accelerated particle.
e = Charge on accelerated particle.
r = Radius of orbit of accelerated particle.
H = Magnetic field strength.
c = Velocity of light.
ω_p = Angular velocity of accelerated particle.

(1) INTRODUCTION

The fixed-frequency cyclotron is one of the most versatile of all particle accelerators. The presence of such a machine in a hospital will make possible the use of a large number of techniques which have not hitherto been available to therapy and medical research.

The machine is designed to produce external beams of about 200 μA of 15 MeV deuterons and about 50 μA of 30 MeV α-particles. These charged-particle radiations are useful in the fields of radiobiology and radiation chemistry, where large beam currents are frequently required at an energy great enough to secure adequate penetration into the target.

Bombardment of a beryllium target by the deuteron beam provides a high-flux source of high-energy neutrons, which, because they are uncharged, have powers of penetration similar to those of X-rays and thus present possibilities of extending the field of radiation therapy.

Both deuteron and α-beams may, of course, be used in the production of radioactive isotopes. Most of the nuclear reactions produced by bombardment in a cyclotron result in the production of an isotope which is chemically different from the target material from which it is made. Thus carrier-free isotopes and high specific activities are available. In this respect a cyclotron may be regarded as complementing a nuclear reactor. The nearness of the source makes possible the use of fast-decaying isotopes, which present obvious advantages for medical work.

The use of an internal target as a neutron source and for isotope production carries with it severe limitations. Maintenance is hampered by the activity induced by the neutron flux in the dee-box and electrode system, and by active surface contamination if the target should start to evaporate through accidental overheating. Experimental effort has therefore been concentrated on raising external beam current, while design has been directed towards reliability, ease of maintenance and rapid fault location and repair.

The authors are with the Medical Research Council.

(2) FUNDAMENTALS

(2.1) Basic Principles

Fig. 1 is a schematic view of the cyclotron showing how acceleration is effected. N and S are the two poles of an electro-

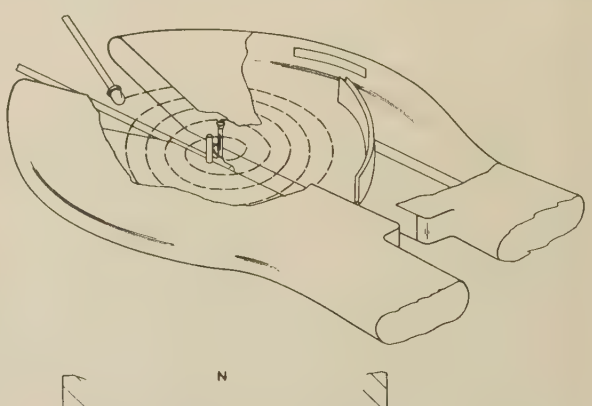


Fig. 1.—Schematic view of cyclotron electrodes showing principle of particle acceleration.

magnet between which is a vacuum-tight dee-box with soft-iron lids. In the dee-box are two hollow D-shaped electrodes, supported on stems extending horizontally into a larger vessel on one side of the magnet. The ion source is situated at the centre of the electrodes and faces one of the dees.

Consider a moment of time when the dee facing the ion-source slot is negatively and the other dee positively charged. Ions will be accelerated from the source into the dee, where they will be in an electric-field-free region. In this region, because of the magnetic field, they will move in a circular path until they again reach the gap between the two dees. If now the polarity of the dees is reversed the ions will be accelerated across the gap; the process will be repeated inside the second dee, and so on. With each increment of energy the radius of the path of the ions

will be increased, so that the path described will be spiral. The angular velocity of a charged particle, moving freely in a magnetic field, is dependent only upon the mass and charge of the particle and the magnetic field strength.

Thus, equating centrifugal and centripetal forces,

$$\frac{mv^2}{r} = Hev \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

From which

$$\omega_p = \frac{v}{r} = \frac{He}{m} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In the above equations m is the relativistic mass given by the equation

$$m = \frac{m_0}{\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2}} \quad . \quad . \quad . \quad . \quad (3)$$

whence

$$\omega_p = \frac{He}{m_0}\left[1 - \left(\frac{v}{c}\right)^2\right]^{1/2} \quad . \quad . \quad . \quad . \quad (4)$$

In a constant magnetic field, and at non-relativistic velocities, the angular velocity of the particle is constant. For deuterons in a field of 15 000 oersteds the resonant frequency is 11·4 Mc/s.

Rearranging eqn. (1), an approximate expression is obtained for the particle kinetic energy,

$$\frac{1}{2}mv^2 = \frac{1}{2}(Hr)^2\frac{e^2}{m} \quad . \quad . \quad . \quad . \quad . \quad (5)$$

In order to keep the particle in the plane of the dees, it is necessary to introduce vertical focusing. This is done by allowing the magnetic field strength to decrease with increasing radius, thus causing the magnetic lines of force to bow outwards. Applying the left-hand rule it can be seen that a particle displaced from the central plane experiences a correcting force.

This weakening of the field, combined with the relativity effect, constitutes the principal limit to output energy attainable. As the velocity of the particle increases, it can be seen from eqn. (4) that its angular velocity decreases and that this effect is enhanced owing to the falling value of H as the radius increases.

Phase excursion is therefore an increasing function of the number of revolutions required to attain a given energy, i.e. it is an inverse function of the accelerating voltage. In practice, the magnetic field strength is increased at the centre of the machine above that required for resonance, so that the particle gains phase during the early part of the period of acceleration and loses phase in the latter part.

The problem of extraction is a complex one and will be discussed more fully in a later paper.

(2.2) Choice of Principal Parameters

Owing to the difficulties inherent in making exact measurements on a working machine, precise information to aid the designer is difficult to obtain. There is, however, abundant experience in operation and design, and from this it was found possible to obtain rough quantitative figures as a guide to ultimate performance.

Since the output energy of a cyclotron is largely determined by the magnet pole-face diameter, and since the magnet is the largest and most costly single item of equipment, there is a rough correlation between cost and energy. On the basis of cost it was decided that the M.R.C. machine should be in the 45 in class, from which an output of 15 MeV 100 μA external-beam deuterons would normally be expected.

The magnet yoke was designed assuming tapered poles, which would give a uniform flux density throughout the pole, except at the root. The final particle radius, however, is determined by the possible magnetic field shape, which in turn is determined by the saturation of the pole edges. It is found in practice that by reducing the field strength and increasing the pole-face diameter the final particle radius is more than proportionately increased, i.e. the product Hr and hence the final particle energy are increased. Further, owing to the reduction of magnetic field strength and the consequent sharpness of the fall-off produced near the edges of the poles, a smaller voltage gradient is required in the deflector. It is not easy to compute the optimum value of pole-face diameter, which was selected as 50 in.

The gap between the lids was chosen as 6 in; this choice was based largely on the Massachusetts Institute of Technology 45 in machine, and on scaling-down from the 60 in machines. From data[1] available from the United States it appeared that up to 100 kV dee-to-earth the maximum working voltage gradient was of the order of 60 kV/in.

Deep dees have the advantage of mechanical rigidity and large beam currents, but, of course, reduce clearance and therefore accelerating voltage. A final clearance of $1\frac{5}{8}$ in was chosen except at the shims, with an internal clearance of $1\frac{3}{4}$ in. The M.I.T. cyclotron works at 70 kV, and phase-shift calculations for the M.R.C. machine showed that 70–75 kV should suffice.

As explained above, the necessary radial fall-off in magnetic field strength provides one of the limitations to output energy, so that the minimum gradient that will give adequate vertical focusing is desirable. Since no experimental evidence appeared to exist as to how small this might be, the limit imposed was the practical one of maintaining azimuthal uniformity. A logarithmic fall-off, giving a constant focusing force, was chosen with an index value of 0·006.

(3) THE MAGNET

(3.1) Principal Dimensions

The magnet proportions are similar to those of the cyclotron at the Atomic Energy Research Establishment, Harwell,[2] and the yoke and pole dimensions were chosen to give a field of 18 500 oersteds in a working gap of $7\frac{1}{2}$ in, assuming poles tapering to 45 in at the face.

The dimensions of the magnet are given in Fig. 2. The yoke

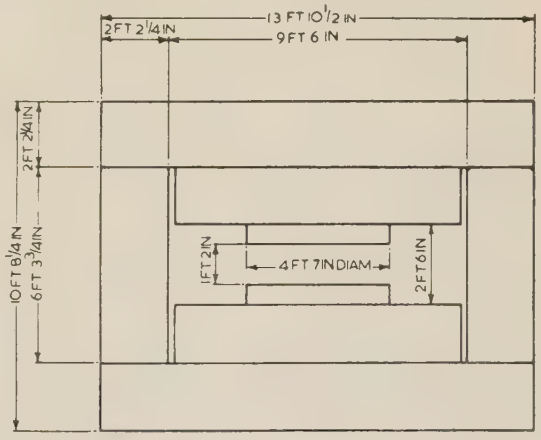


Fig. 2.—Magnet dimensions.

The yoke members are 4 ft 9 in deep, having the same dimension as the pole root.

consists of four solid forgings, and each pole consists of two cylinders 57 in in diameter. The steel has an 0·1% carbon content, and the pair of cylinders which form the pole tips were horizontally cast to ensure homogeneity. The coils are wound from 1·15 in × 0·75 in section aluminium, with an internal bore

of 0·4 in diameter through which distilled water passes. Each coil is built up from eight double pancakes wound so that all joints are on the outside. The pancakes have their electrical paths connected in series and their water paths in parallel, the connections to the water header being made by reinforced synthetic rubber tube. This material was chosen because, ultimately, the cooling water may contain 1% potassium dichromate to restrict possible electrolytic action, and tests showed that many materials, including natural rubber, would deteriorate in this solution.

(3.2) Magnet Excitation and Stabilization

The windings are designed to give $4{\cdot}6 \times 10^5$ ampere-turns. Each coil has 408 turns, and thus full load rating is 560 amp. This is obtained from a 155 kW air-cooled motor-generator. The generator has two field windings; the main winding of 16000 ampere-turns provides the excitation, and the auxiliary winding of 5600 ampere-turns is connected to the output of an amplidyne motor-generator. The latter is the last stage in a circuit for stabilizing the magnet current to one part in 10^4. The circuit is shown schematically in Fig. 3. For the reference signal 1 volt

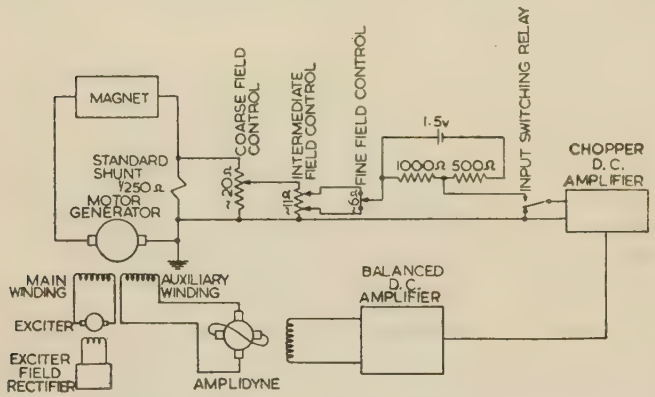


Fig. 3.—Magnetic field stabilization circuit.

was preferred to the more usual 0·1 volt, to reduce the difficulty of handling small signals at the input stage. This was particularly necessary because standard telephone relay contacts are used in the input circuit to protect the dry cell and d.c. amplifiers when the circuit is not stabilizing, and these contacts can contribute small spurious signals. The shunt dissipates about 750 watts at full load and is water-cooled. Operational experience of the stabilization system shows that, provided that the relay contacts are cleaned once a month, the system will operate with the full loop gain of 240 without trouble. A commercial 'chopper' type d.c. amplifier has proved completely satisfactory for the first stage of amplification. When the magnet is switched on the field is about 0·5% high; this settles down to a steady value in about 15 min, after which the drift is less than 0·05%.

(3.3) Magnet Pole Tips and Dee-Box Lids

For the shimming gap to be effective the dee-box lids must be thin; if they are too thin, however, they will distort under the vacuum and magnetic forces. As a compromise a thickness of 3 in was chosen, after tests on a model had shown that the vacuum forces would dome each lid by no more than 0·005 in.

The 1 in shimming gaps, which are rather larger than usual, enable shimming adjustments to be made easily, and allow sufficient space for the installation of proton-resonance magnetic measuring equipment. Tests using a model magnet showed that such large shimming gaps produce no adverse effects on the magnetic field shape.

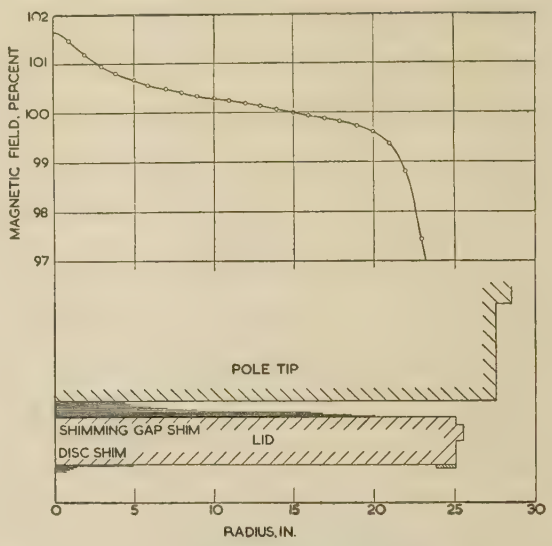


Fig. 4.—Variation of magnetic field strength with radius and configuration used to produce it.

The shimming-gap shims are all $\frac{1}{32}$ in thick. The steps on the disc are about 0·020 in thick.

It will be seen from Fig. 4 that the poles are almost cylindrical. The root is 57 in in diameter and the tip 55 in, a 1 in step occurring $6\frac{1}{4}$ in back from the tip. The sole purpose of this step is to provide a location on the lower pole for the bronze frame upon which the dee-box is supported; the step is repeated on the upper pole for reasons of symmetry. The pole tips overhang the 50 in diameter lids by $2\frac{1}{2}$ in. This arrangement, rather than the more conventional taper, is the result of model-magnet experiments in which the various shapes shown in Fig. 5 were tried. It was

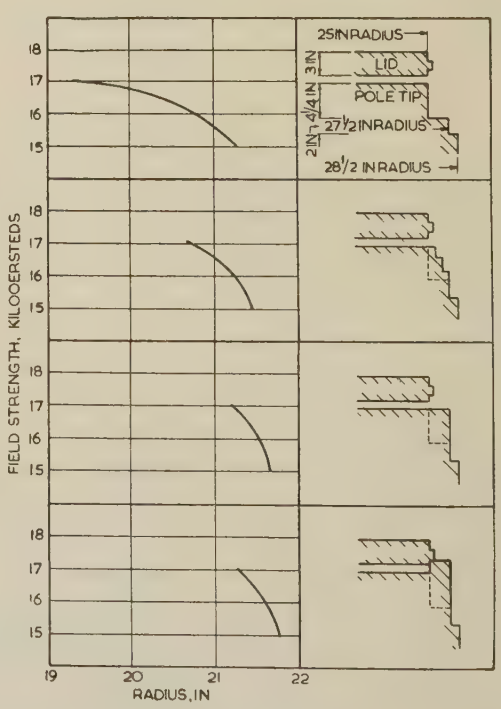


Fig. 5.—Effect of pole shape and field strength on the radius of the last turn.

It is assumed that the last turn occurs at the point where the magnetic field strength is 2·5% below its value at 15 in radius. The measurements were taken on a 2 : 15 scale model, but all dimensions are given in terms of the cyclotron magnet.

found that the overhang gave a slightly larger working radius, a steeper fall-off over the extraction region and less change in field shape with field strength than the taper.

The tolerances set for the magnet gap were: pole and lid faces flat to within 0·002 in; opposite poles and lids concentric to within 0·010 in; parallelism of pole faces to within 0·005 in; parallelism of lids to within 0·002 in. Armco ingot iron was chosen for the lids, which were tested with a supersonic crack detector. Many flaws just below the surface and of about $\frac{1}{4}$ in diameter were located, but they produced no measurable effect on the magnetic field.

The top and bottom of the dee-box consists of square 2 in thick brass plates, joined at the corners by stout brass posts. The plates are bored out to take the iron lids which are fixed in place by clamping rings and sealed by rubber rings. In the absence of magnetic and vacuum forces the whole structure is supported by corner jacks off the bronze frame on the bottom pole, thus giving access to the shimming gaps above and below the box.

The open structure of the dee-box allows the top and bottom plates to flex under the forces of atmospheric pressure and the magnet. With vacuum on alone, the top and bottom plates come together, limited by struts at the middle of the open sides, but, with the magnet on, the plates are drawn apart so that the lids come up against spacers in the shimming gaps.

(3.4) Choice of Magnetic Field Shape

The logarithmic field fall-off has an exponent $n \left(= \frac{r}{H}\frac{dH}{dr}\right)$ of 0·006. This gives the particle a period of vertical oscillation equal to 13 revolutions, so that a particle leaving the central plane at an angle to the horizontal will reach its maximum amplitude in just over three turns. Numerical plots of the first turns show that, while electric forces are invariably strong in the first turn, the magnetic focusing forces may predominate subsequently. Accordingly the logarithmic fall-off was chosen to commence at 2 in, which is about the radius of the first turn.

Three types of deviation from the ideal field shape arise:

In the first type the gradient itself varies. It was decided that the sense of the error, where it was unavoidable, must be such as to make the gradient steeper than necessary, since loss of phase would be less serious than loss of focusing.

The second type of deviation is in the azimuthal uniformity of the field. The mechanics of the particle motion are such that in their excursion through the cyclotron the orbits precess[3] for about one turn about the magnetic centre of the machine. Furthermore, if the particles are to be extracted, their final orbit centres must lie about 1 in off-centre in the direction of the entrance of the deflector channel. While it is true that to achieve extraction it has often been found necessary to introduce small offsets in the magnetic centres, these could not be determined in advance and so it was decided to keep the magnetic centres within 0·125 in of the true centre. This is satisfied, providing that the amplitude of the first harmonic of the azimuthal variation of magnetic field is not numerically greater than one-eighth of its radial gradient. Higher harmonics may be neglected.

The third type of deviation arises when there is vertical asymmetry in the magnetic field. The curvature of the lines of force which gives rise to vertical focusing is so small that at 18 in radius, with $n = 0{\cdot}006$, the centre of curvature of a line of force lies 250 ft away from the machine. If this centre of curvature were to rise only $\frac{3}{4}$ in above the central plane the entire beam would be lost to the dees.

The surface about which particles are focused is often called the median plane; but it will be referred to as the equilibrium plane, to avoid confusion with the geometrically central plane. It should be noted that, whilst this plane cannot be appreciably tilted, its height may vary with radius. It was decided that it would not be allowed to deviate more than $\frac{1}{4}$ in from the central plane. The variation of the equilibrium plane height with radius which was actually achieved is shown in Fig. 6.

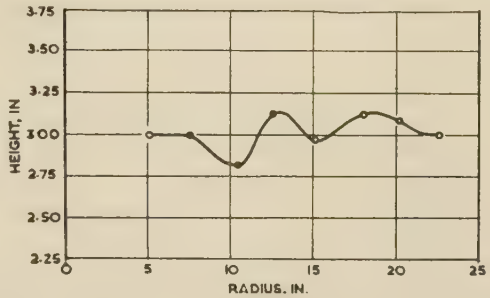


Fig. 6.—Variation of equilibrium plane height with radius.

(3.5) Shimming Procedure

(3.5.1) Main Field.

Based on experience with the model magnet, the following procedure was adopted for producing the desired field. Before the lids were inserted, the field in the 14 in gap between the pole tips was checked for azimuthal uniformity with a tracking coil. This was found to be uniform to within one oersted in 10 000, which was more than adequate. An azimuthal plot was taken with the lids in position and the shimming-gap spacers were altered until the plot was again uniform to one oersted.

The effect of shimming on the field of the model magnet is shown in Fig. 7. The final field of the cyclotron magnet was built up by working from the outside inwards, whilst holding the

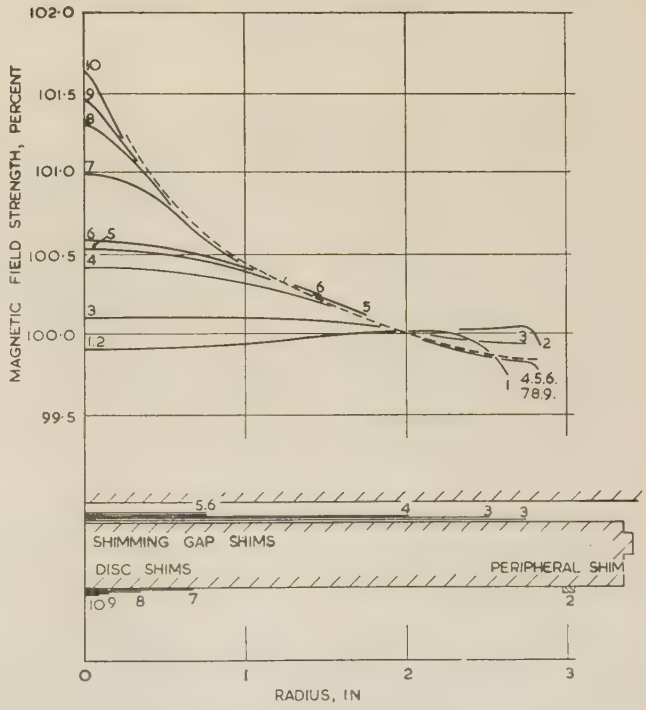


Fig. 7.—Model magnet plot showing build-up of shims.

The graph shows the percentage change in magnetic field strength as a function of radius as shims numbered 2–10 are added. The required logarithmic field is shown dashed. Model magnet dimensions are given. The shimming-gap shims are 0·006 in thick and the vacuum-gap shims are 0·002 in thick.

field strength constant at 15 in radius, where the particles are in resonance with the radio frequency, by means of a second proton-resonance coil. This method of working was found to be essential as the larger-diameter shims contribute to the magnetic field produced by the smaller ones.

The main-field gradient was obtained by building a pyramid of circular shims each $\frac{1}{32}$ in thick. These were sandwiched between pieces of plywood and screwed together for convenience in handling, the whole pack being moved in one unit whenever a new shim was added (see Fig. 4). To take out a pack it is necessary to remove two of the spacers from the shimming gap. This is done by evacuating the dee-box, when, in the absence of a magnetic field, the spacers are free.

(3.5.2) **Peripheral Shims.**

Various sizes of Armco ingot iron annuli, covering a quarter of a circle of the top and bottom poles, were tried until a proton-resonance check over the field fall-off region showed an arrangement which gave the maximum exit-slit radius, without producing a bump in the field. Peripheral shims, $1\frac{1}{4}$ in $\times$ $\frac{1}{4}$ in, covering the full circle were then made and screwed into place.

(3.5.3) **Fields at Small Radii.**

At small radii the shimming gap is ineffective, so the first 7 in of fall-off is produced by a pair of 10 in diameter iron discs, recessed $\frac{1}{4}$ in into the lids. These discs were made initially oversize and then machined down on the face until the required field was obtained by trial and error.

The discs are cut away on the underside leaving a ring $\frac{1}{2}$ in wide by 0·020 in proud on the periphery. This ring rests against the recess in the pole face, which is $\frac{1}{8}$ in larger in diameter than the disc. This arrangement causes the disc to dome slightly, and hence prevents tilting by providing an alternative position of stable equilibrium. A small screw at the centre retains each disc in position when the magnetic field is switched off.

It was found necessary to correct azimuthal uniformity, after the required field shape had been obtained, by a small sideways movement of the shimming packs.

(3.5.4) **Location of the Equilibrium Plane.**

The position of the equilibrium orbits was next determined using the floating-wire method.[4] Loops varying in radius by $2\frac{1}{2}$ in from 5 in to $22\frac{1}{2}$ in were used. These showed the equilibrium plane to be as much as $\frac{3}{4}$ in high in the region of the central cone, and $\frac{1}{2}$ in low at $12\frac{1}{2}$ in radius. The removal of metal from the underside of the bottom disc and of one shim in the top pack gave the necessary correction, and the equilibrium plane was then central to within $\frac{1}{4}$ in.

A check of the azimuthal uniformity and the radial fall-off showed these to be unchanged within the limits allowed.

4) THE RESONATOR AND ELECTRODE SYSTEM

(4.1) The Resonator

The electrode system, illustrated in Fig. 8, is in the form of a screened pair, since this gives a high Q-factor, facilitates the separation of undesired modes of operation and gives a better pumping speed.

The resonator was analysed as a system with the dees regarded as a short length of low-impedance transmission line, open-circuited and loading a longer length of short-circuited line of higher impedance. This gave the resonant frequency and power requirements, which were confirmed by measurements on a 1 : 10 scale model. The measured Q-factor of the resonator, 4 500, proved to be about 70% of the calculated value. Reports of other cyclotrons indicate that this discrepancy is usual.

(4.2) Dees

The factors governing the choice of the principal dimensions of the dees have been discussed in Section 2.2.

In order to reduce, as far as possible, the amount of induced radioactivity, the inside surfaces of the dees are covered with carbon blocks (Figs. 9 and 10) in such a way that the beam never sees the copper of the dees. It is common practice to provide feelers on the dee facing the ion-source slit in order to increase the electric field gradient and so increase the accelerated beam current. A vertical feeler, consisting of a carbon cylinder of $\frac{1}{2}$ in diameter, wedged between the top and bottom edges of the dee is used, as this also reduces vertical defocusing (Fig. 9).

The dees (Fig. 10) are mirror images of each other and each is made in two halves split along a horizontal plane. They are stiffened internally by an aluminium-alloy spine and an I-beam structure at the junction with the taper section. In addition to splitting the dees for internal access, a large removable plate is provided over the deflector and septum assembly for maintenance purposes. All screwed joints carrying large currents on the dees are provided with cooling tubes. The joint carrying the highest mechanical stress is that between the cylindrical brass stems and the taper section of the dee. A steel casting is keyed into each part of the joint and the two castings are bolted together.

The dees are supported on 8 in brass tubes held in gimbals in the vacuum take-off box. The ends of the tubes project through vacuum seals at the back of the box. Operating on each of these projections are two motor thrusters and a return spring, by means of which the vertical and horizontal position of the dees may be remotely controlled.

The conducting surfaces of the dee stems are composed of 24 T-section copper cooling tubes, with the cross-bars of the T's forming a 9 in diameter cylinder, the whole being clamped around the brass-tube structural member. These are assembled as two semi-cylindrical shells, brazed to copper half-collars at the dee end. The connection between these collars and the dee surface, at the junction of the parallel and taper sections, is by means of screws having hardened beryllium copper knife-edge washers which bridge the gap.

(4.3) Short-Circuit

A general view of the short-circuit is shown in Fig. 11. The electrical length of the dee stems is determined by the position of a structure short-circuiting the stems to each other and to the conducting liner of the parallel section. This short-circuit is movable in order to accommodate manufacturing errors throughout the resonator and provide for the possibility of subsequent energy changes.

The use of a screened pair means that the forces required to make contact are not balanced between the screen and the dee stem; the contacts were therefore made with local reaction members. This design relieves the conducting plane of the short-circuit of any mechanical requirements, thus permitting it to be designed for flexibility and good pumping speed.

Contact is made to the T-section tubes which form the dee-stem conductors, and to similar tubes brazed to the screen. Beryllium-copper knife edges are forced into the copper by a stiff leaf spring, set up by a lever-operated cam, the reaction from which is transferred to the lower side of the T.

The conductor between the contact blocks (Fig. 12) consists of two $\frac{3}{16}$ in copper tubes, curved for flexibility, which bring water cooling close to the contact edges. The current is evenly distributed among these tubes by making them follow the theoretical current paths in a semi-infinite plane.

(4.4) Screen

The dees and dee stems are screened by a copper liner inside the parallel section and taper section of the vacuum vessel, and

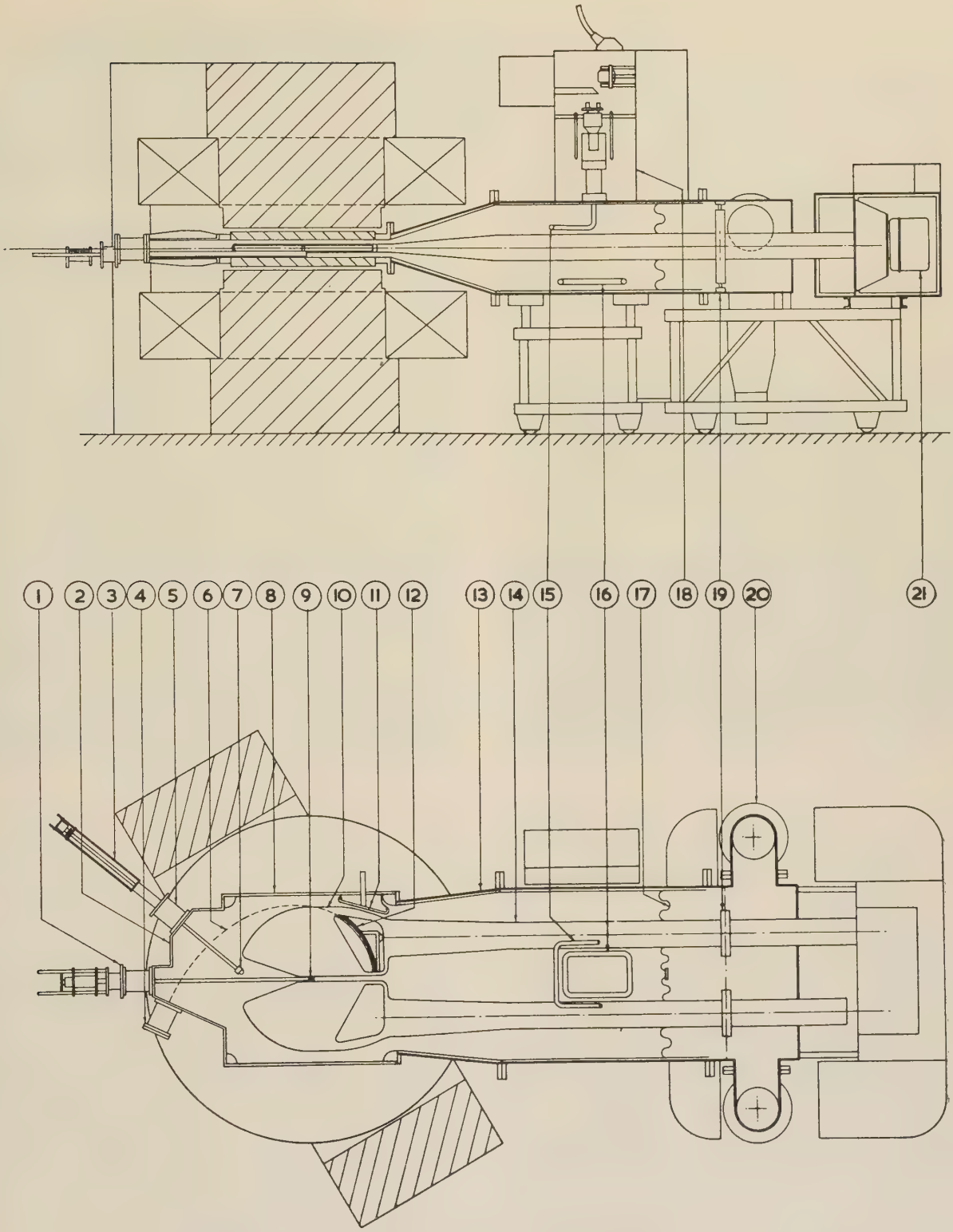


Fig. 8.—Part sectional plan and elevation of cyclotron assembly.

1. Ion-source adjusting mechanism and gate valve.
2. Dee-box extension.
3. Rotating-target drive mechanism.
4. External-target gate valve.
5. Rotating-target gate valve.
6. External beam path.
7. External target.
8. Dee-box.
9. Ion source.
10. East dee.
11. Trimming condenser.
12. Deflector and septum.
13. Copper-lined vacuum vessel.
14. Dee stem.
15. Coupling loop.
16. Tuning loop.
17. Short-circuit.
18. Power amplifier.
19. Gymbal.
20. Diffusion pump.
21. Deflector terminal.

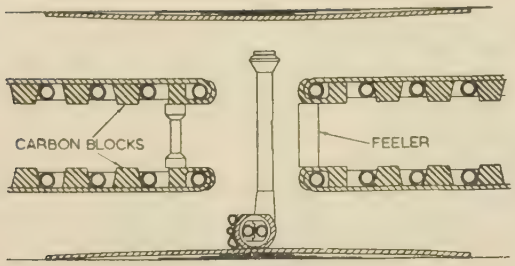


Fig. 9.—Transverse vertical section through dee interface.

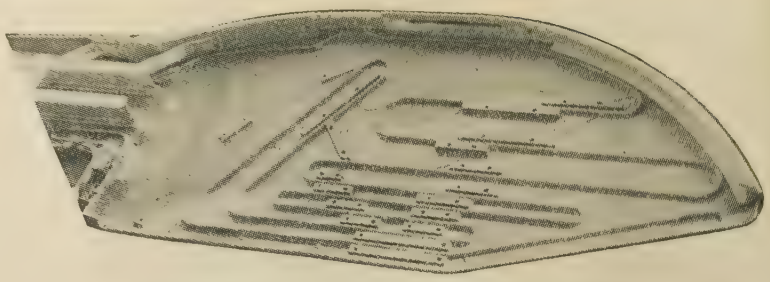


Fig. 10.—Bottom half of split dee showing protective carbon blocks, aluminium spine and water-cooling tubes.

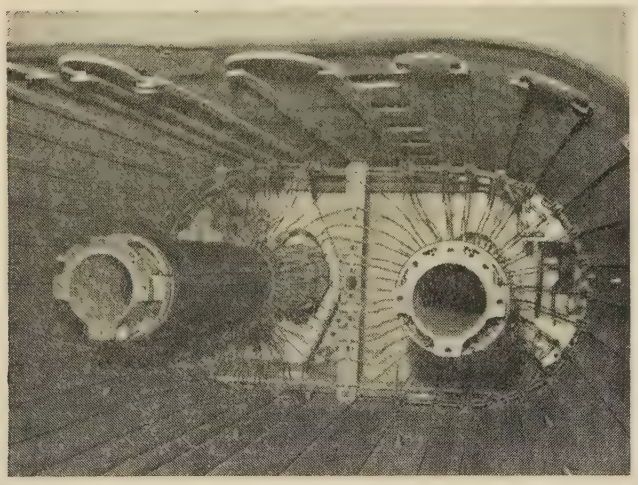


Fig. 11.—Short-circuit viewed from dees end during assembly, prior to fitting of dees.

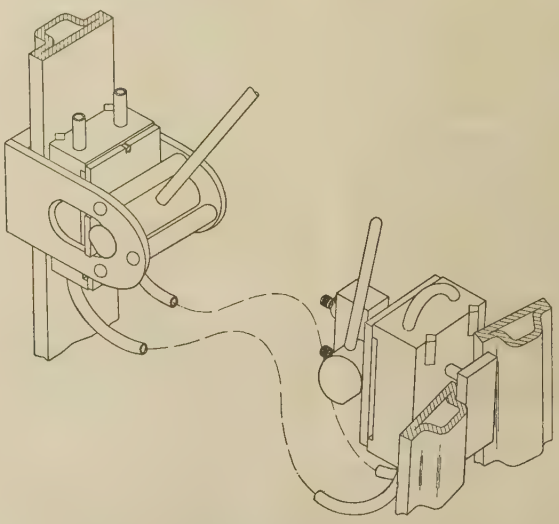


Fig. 12.—One pair of short-circuit conductors showing outer and inner contact blocks.

by copper plating on the dee-box interior and lids. Contact is made between the faces of the flanges on these sections by soft copper foil wrapped around rectangular-section rubber strips. The rubber is supported on a brass strip screwed to the flange.

(4.5) Tuning Loop

In order to maintain resonance of the cyclotron despite thermal and beam loading changes, some form of variable reactance is required. This takes the form of a horizontal water-cooled short-circuited turn, measuring about 1 ft 6 in × 2 ft 6 in, which may be raised between the dee stems near the short-circuit, providing about 2% total change in frequency. A single trimmer condenser is used on one dee to balance the dee voltages.[5] The setting of the trimmer remains constant during operation, but the tuning loop is continually adjusted by automatic control. At the extreme position of the loop it may be called upon to dissipate 10–15% of the total r.f. power consumed.

(4.6) Coupling Loop

The r.f. output from the power amplifier is coupled to the dee stems by a single loop connected through contact plates on glass insulators, capacitance-coupled to the anodes of the valves. The voltage ratio between the valve anodes and the dees, set now to about 1 : 9, is varied by vertical adjustment of the coupling loop, provided by telescopic tubes. Electrical and thermal contact at the sliding joint is provided by four small-area contacts, the heads of $\frac{1}{16}$ in diameter brass rivets, firmly pressed to the inner tube. These contacts are under relatively low pressure, but the current is small (10 amp).

(4.7) Dee Voltmeters

The dee voltage is measured by a capacitance-potentiometer diode peak voltmeter. The probe head is flush with the copper liner of the taper section, facing an approximately flat part of the dee-stem taper. The probe head is extended as a thick conductor through vacuum seals to the atmosphere to promote heat loss. Field emission currents may raise the temperature of the head considerably, and can cause spurious current flow through the meter and heating of the diode filament. The heat loss is increased by the provision of a radiating terminal disc, which acts also as a support for components. The voltmeter is subject to small errors due to vertical movement of the dees.

(5) THE RADIO-FREQUENCY SYSTEM

(5.1) General

For the r.f. generator a power amplifier driven by a master oscillator was chosen in preference to a self-oscillator. A driven amplifier avoids 'multi-pactoring' difficulties; it also permits power limitation by drive reduction during voltage breakdown and allows separate testing of the r.f. equipment before completion of the cyclotron.

(5.2) Power Amplifier

The power amplifier (Fig. 13) uses two water-cooled BW165 valves in a conventional neutralized earthed-filament push-pull circuit. This arrangement is capable of an output of 85 kW when running from a 10 kV glass-bulb mercury-arc rectifier at its

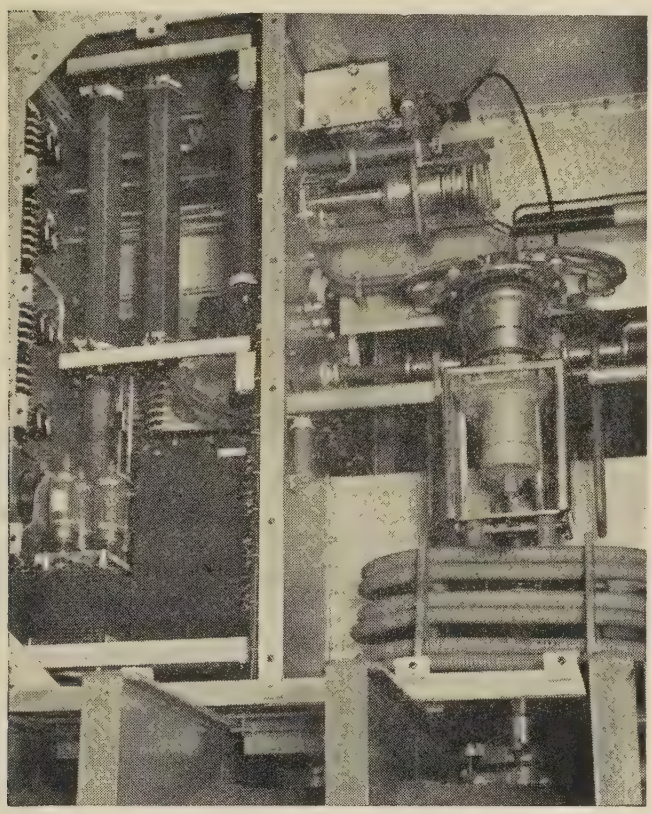


Fig. 13.—One side of symmetrical push-pull power amplifier.

two-hour rating of 12 amp, but voltage breakdown in the cyclotron limits the output to about 50 kW, plus beam loading.

Grid bias is automatic from 1 000-ohm grid resistors, with 300 volts of safety bias from separate rectifiers to permit full-range drive control. Hum level is reduced to less than 1% by generous h.t. smoothing and a 6-phase rectified filament supply.

The only resonant impedance in the anode circuit is the coupled impedance of the cyclotron, which appears as 1 000 ohms at resonance, in series with the coupling loop reactance of about 50 ohms. The grids are coupled by a π-network from two 84-ohm coaxial cables leading from the driver unit.

Parasitic oscillations were found on test. A parallel mode at about 1 Mc/s was eliminated by increasing the resonant frequency of the grid-decoupling circuit. A more violent oscillation at 100 Mc/s, due to cavity resonance of the cabinet, could be eliminated only by inserting a series-resonant circuit between grid and filament.

(5.3) Driver

The driver was modified from a commercial telegraph transmitter, consisting of four identical r.f. units. Two of these units are in use as the halves of a push-pull amplifier from a common master oscillator, with two as standby. The outputs are inductively coupled to the coaxial lines feeding the amplifier, with a Faraday screen to reduce harmonic radiation. The master oscillator is crystal controlled at 11·286 Mc/s.

(5.4) Automatic Tuning

In order to accommodate changes of cyclotron tuning due to temperature changes, more rapid smaller variations caused by the reactive loading of the beam, gas discharges and breakdown, an automatic tuning device is installed which operates the tuning control to maintain 180° phase difference between grids and anodes of the power amplifier.

Small capacitors couple the grids and anodes to 100-ohm terminated lines, which are inductively coupled, with screened coupling coils, to tuned circuits. The power-amplifier grids feed the anodes of a double triode in phase opposition, and the anodes, after two stages of limiting amplification and a goniometer phase shifter, feed the grids in parallel. The d.c. component of the anode voltage, which indicates deviation from 90° phase relationship, is amplified and superimposed upon a triangular wave, feeding two relays which control the tuning. The period of the relay response is thus made proportional to the amount of detuning.

In order to avoid spurious detuning under trip conditions, a relay is provided to render the apparatus inoperative until drive is present at maximum level.

(6) THE ION SOURCE

The construction of the ion source is shown in Fig. 14, and its position relative to the dees in Fig. 9. In the presence of the strong vertical magnetic field, the arc current is confined to

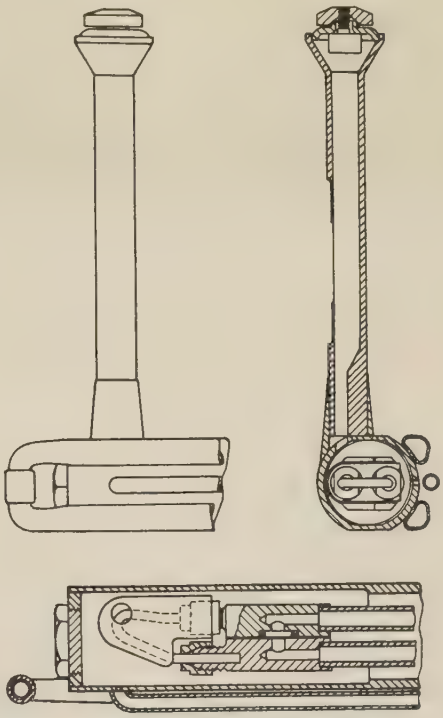


Fig. 14.—Ion source.

the vertical direction. A $\frac{1}{8}$ in diameter hole pierces the arc chamber immediately above the hottest part of the filament. The ionizing electrons and the gas pass through this hole into the carbon chimney. The electrons ionize the gas, and the electric field from the dees draws ions from the plasma through a slot in the side of the chimney. At the top of this chimney is a quartz-insulated electrode which, owing to the mechanism of electron dumping, closely follows the potential of the filament. Thus a type of Penning discharge is set up, which greatly increases

the efficiency of ionization. This design of chimney is similar to that adopted by Oak Ridge National Laboratories.[6] Attention has been paid to compactness, in order to reduce the disturbing effect of the tube on the electric field. To this end the chimney is offset to bring the ion column as close to the dee as possible without bringing the arc chamber under the dee. This not only reduces the disturbance to the electric field but also increases flash-over voltage, which, in the magnetic field, is more sensitive to reductions in vertical clearances.

The conductors carrying the filament current are single water tubes with an insulating water link across the filament blocks; the link consisting of a silicone-rubber ring which has to maintain a vacuum-tight joint in the presence of temperatures which destroy normal rubber.

The filament consists of a 'hairpin' of tungsten, of 0·1 in diameter, which requires about 250 amp to raise it to emitting temperature. Contact to the ends of the filament is by collets on the filament blocks. Filaments of this design last between 20 and 30 hours. The filament is supplied from a 6-volt 500 amp d.c. generator.

The arc supply is from a 600-volt 5 amp glass-bulb mercury-arc rectifier, through an adjustable dropping resistor of about 175 ohms which reduces the arc voltage to between 150 and 250 volts at the normal arc current of 2–3 amp.

Owing to variable emission, the arc tends to wander over the filament. To avoid this and to ensure a stable current in the chimney, a molybdenum shield encloses the whole of the filament except the region immediately under the $\frac{1}{8}$ in hole. It is found that this shield doubles the column current. Approximate current distributions in the source are as follows: total cathode current, 2 amp; current in chimney, 100 mA; total ion current drawn from chimney slot, 10 mA.

There is provision for adjustment of ion-source position both vertically and horizontally and for twisting about a horizontal axis. The filament may also be moved in and out to adjust for the best position under the chimney.

(7) DEFLECTOR

It is desirable that the beam should travel as short a distance as possible in the fringing field, to avoid excessive horizontal defocusing, and that it should leave the cyclotron approximately normal to the long axis of the magnet, to give maximum space round the beam for experimental equipment. To improve beam direction in the latter respect singly flared dees are used, permitting the deflector channel to start near the dee interface.

The sides of the deflector channel are at present made of carbon. These glow brightly when the beam strikes them and enable the channel to be aligned by visual observation. The deflector side of the channel can be seen through a series of $\frac{1}{8}$ in holes drilled through the septum. There is considerable emission from the hot carbon when a 125 μA beam strikes the channel entrance, but this does not cause voltage breakdown. When the larger deuterium beam is circulated, the carbon at the entrance to the channel will be replaced with copper.

The deflector may be moved so as to vary the channel width at either end independently while the machine is running. The septum may be moved without breaking the vacuum by screw adjustments made through the dee aperture.

The stem of the deflector is supported on a p.t.f.e. insulator with re-entrant sputtering shields situated in the tapering section of the dee stem, and is also supported at the glass insulator at the outer end of the dee stem. The glass insulator is 9 in in diameter and 12 in long, and contains sputtering shields to protect the glass. The stem of the deflector consists of an aluminium tube containing three smaller tubes, of which two carry cooling water to the deflector itself and one acts as a push-pull operating rod for adjusting the width of the deflector channel at the outlet end.

The electrical supply to the deflector is by a two-valve voltage-doubling circuit with an effective reservoir capacity of 0·05 μF, an *RC* filter of 10 megohms and 0·01 μF, a current-limiting resistor of 10 megohms and metering chain of 100 megohms. Cooling of the deflector is by distilled water through a pair of polythene tubes 65 ft long by $\frac{1}{4}$ in bore wound on a polythene former.

(8) TARGETS

The cyclotron beam may be intercepted either internally or externally. The internal target is inserted in the flared gap between the dees. At this point the beam is highly concentrated, delivering up to 15 kW of energy on to an area of only a few square millimetres. To spread this power over a greater area and to avoid melting, a 2 in diameter spherical target head, rotating at 50–150 r.p.m., is presented at an angle of 45° to the beam. Materials of low melting point or low thermal conductivity are not suitable for use as internal targets. The resistance to voltage breakdown of the dees is a critical factor in satisfactory operation of the machine, and is highly sensitive to the contamination that would result from the overheating of an internal target.

The external target intercepts the beam after it has passed through the deflector channel. Owing to horizontal defocusing in the fringing field the beam usually covers an area of 10–20 cm^2, giving a density of about 1/500 of the internal beam. In order to isolate the target material from the vacuum system, the beam will be made to emerge through a 0·001 in metal window before hitting the target.

The fitting of experimental targets and probes for tracking out the beam has been greatly facilitated by the provision of three removable plates, 5 ft × 8 in, to the sides of the dee-box. This arrangement gives maximum access to the interior and makes it easy to design extension boxes for the external beam.

(9) THE VACUUM SYSTEM

The acceleration of the particles is carried out in a high vacuum at a pressure of about 10^{-5} mm Hg. A vacuum-tight steel shell encloses the resonator and extends beyond it to provide mechanical support for the dee stems and a manifold for the two 16 in diffusion pumps. These pumps are fitted with baffles refrigerated to −40° C in addition to the usual water-cooled baffles to reduce backstreaming of oil into the vessel.

The volume of the vessel is approximately 3 000 litres and the total pumping speed of the two pumps is about 3 000 litres/sec, giving rapid removal of gas.

The diffusion pumps are backed by two rotary pumps of 200 ft^3/min capacity each. Two small rotary pumps are provided to maintain vacuum in the diffusion pumps when the latter are isolated, leaving the larger pumps free to pump the vessel.

All valves and pumps in the vacuum system are automatically operated by a relay system which takes information from a series of vacuum gauges and valve-position detectors and is arranged to close down and shut off such parts of the system as are in danger owing to a fault. It is also possible to pump down the vessel to full vacuum simply by initiating the process and leaving the relay system to provide sequential control.

For automatic control of the vacuum-pumping system it is necessary to detect a pressure transition in the region of 200 μ Hg, below which the diffusion pumps may be safely operated, and also a transition at 10^{-4} mm Hg, below which the hot-cathode ionization gauge may run without damage. A straightforward spark-gap, the current from which is arranged to fire a cold-

cathode triode, provides the sensitive element for the 200 μ Hg point. The spark-gap current drops sharply in the region of 200 μ Hg and so extinguishes the triode and releases its anode relay. Protection against maloperation due to failure of supply is by means of a relay detecting the spark-gap and triode supply. The spark-gap current also drops at about 200 mm Hg, but extinction at this pressure is distinguished from that at 200 μ Hg by a bellows-type pressure detector set to operate below 200 mm Hg. The combination of information from the bellows and spark gauges can thus give unique indication of transition of the 200 μ Hg point.

Pressure measurement is by means of Pirani gauge between 1 mm and 1 μ Hg, by Penning gauge between 1 μ and 0·1 μ Hg, and by hot-cathode ionization gauge between 0·1 μ and 10^{-5} mm Hg.

The ionization gauge is fitted with an electronic voltmeter having very low output impedance, so that a variable number of parallel indicating instruments with their own series multipliers may be used at different parts of the plant without affecting accuracy of reading.

(10) COOLING

The cooling system for the cyclotron has been divided into three independent sections. The magnet cooling was separated from the rest of the machine because of the special precautions taken to prevent corrosion of the aluminium coils. The vacuum-pump cooling is run continuously and represents a steady load; it was therefore separated from the tank-cooling circuit, which includes the r.f. electrodes, dee-box lids, deflector, targets and amplifier valves.

Each cooling circuit comprises an evaporative cooler with motor-driven fan, circulating pumps in duplicate for the cooling (mains) water, a heat exchanger, circulating pumps in duplicate for the distilled water, a distilled-water header tank, and the load.

The temperature of the distilled water entering the magnet is maintained at 86° F, by means of by-pass valves round the heat exchanger controlled by a pneumatic thermostat on the inlet water. The tank-circuit distilled-water temperature is controlled by a thermostat in the return water; the control therefore tends to keep the electrode system at a steady temperature, even when no power is being dissipated. This arrangement reduces the detuning which occurs as a result of distortion of the dees due to rapid cooling whenever the r.f. power is taken off. The vacuum-pump circuit has a bleed-in connection, arranged to feed mains water into the system when the temperature of the return water rises above 86° F.

All cooling paths are provided with flow relays, resistance thermometers and thermostats, which are used for interlocking or for indication in the control room. By appropriate interlocks, excess temperature in the magnet, vacuum pump or r.f. amplifier valve cooling circuits shuts down the equipment concerned. Excess temperature in the tank-cooling paths gives a warning and indication in the control room, leaving the operator free to stop or continue to run as appropriate.

(11) POWER SUPPLIES

The power supplies for the cyclotron building are delivered at 415 volts a.c. through two parallel feeders to a switchboard in a room behind the cyclotron control board.

Busbar trunking is provided in the ground- and first-floor plant rooms (200 amp), and on the west wall of the cyclotron chamber (100 amp). All auxiliary apparatus is fed from these busbars through fixed-type fuseboxes. The busbar trunking has enabled additional equipment to be installed with the minimum of difficulty in providing power supplies.

(12) CONTROL SYSTEM[7]

(12.1) General

The control system brings the control and indication of all sections of the plant together in a central control room, provides the necessary safety and operational interlocks between various sections of the plant, and provides protection for personnel, both from radiation from the cyclotron and also from danger arising when working on the auxiliary plant.

Telephone-type equipment is used for the control system, with relays operating at 24 volts d.c. To enable faulty equipment to be changed quickly, and to avoid the extensive testing which would be required if wiring had to be detached from the terminals, many components have been arranged to plug into circuit. Telephone relays are grouped on 12-way mounting plates which plug into cradles having 96 jack connections. Individual relays are mounted on a bracket carrying a 10-way plug; the associated socket carrying the fixed wiring is flanked by two springs which retain the relay. Chassis wiring terminates on plugs connecting with sockets fitted on cradles which are attached to the chassis runners in standard racks (Fig. 15). Spares are

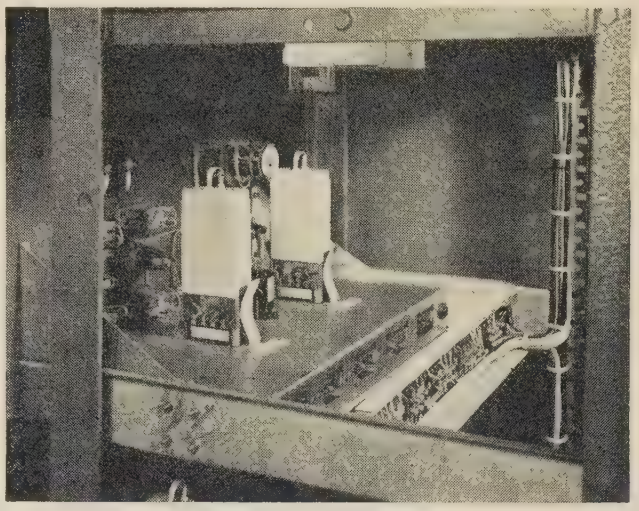


Fig. 15.—Chassis arranged to plug in.

carried for all relay mounting plates and chassis so that the cyclotron can be kept in operation whilst repair or maintenance is carried out in the shop.

A numbering system has been used for terminals which by means of a code indicates geographical location. For example the terminal designation E9IIA6 shows that the terminal is in zone E (ground-floor plant-room south end); equipment 9 (stabilization rack); unit II (excitation-power chassis); terminal group A; and is terminal No. 6.

By incorporating this code in the circuit diagrams, the necessity for wiring diagrams for fault location has been eliminated.

(12.2) Control Room

The control room contains a control board, which forms one wall of the room, and a control desk.

Individual controls are provided on the board for each item of equipment, which can thus be started and stopped independently of all other equipment, subject only to sequential and safety interlocks. The circuits controlled from the board are arranged so that they can be ganged together for operation from the control desk. The controls on the desk are confined to on and off pushbuttons for the three gangs, and pushbuttons for setting operating values for ion-source filament generator excitation, ion-source gas-flow rate, deflector h.t. voltage, r.f. amplifier

h.t. supply and r.f. tuning loop position. Meters on the desk are restricted to ion-source filament voltage, filament current, arc current and gas flow; ion gauge; deflector voltage; west- and east-dee peak voltage; and a large central meter for beam monitoring.

(12.3) Local Control Points

Local controls are provided for all sections of the plant adjacent to the equipment itself. Control is transferred from the control room to local controls by key-operated switches, thus ensuring the safety of personnel working on the plant.

(12.4) Radiation Protection

Protection of personnel who enter the cyclotron chamber has been achieved by interlocking circuits between the cyclotron and the chamber door. Experimenters will have individual personal keys which will release the chamber east-door key from the

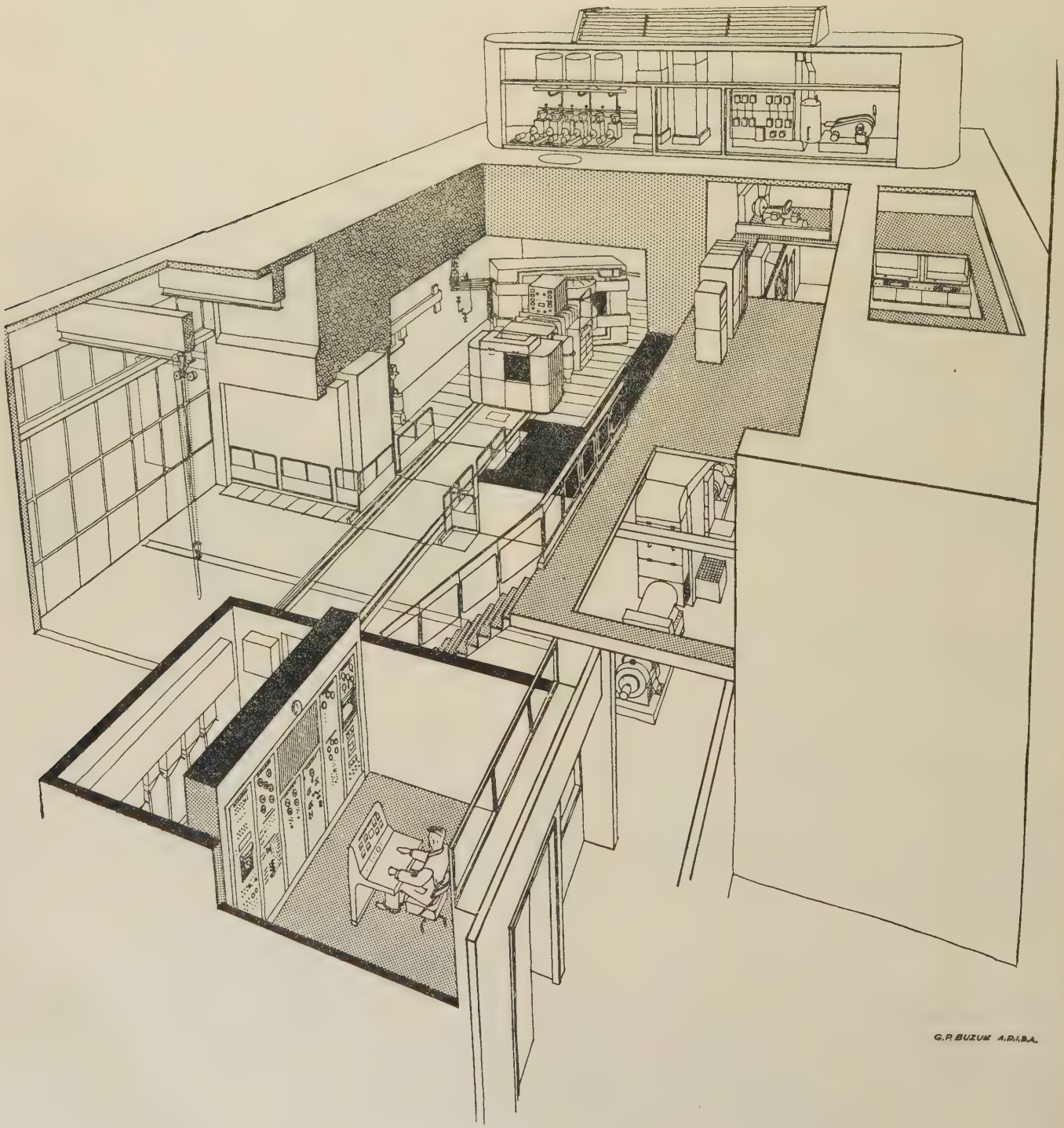


Fig. 16.—Cyclotron suite.

control board. The individual key used and the time of release are marked on an operation recorder.

Responsibility for the safety of personnel in the chamber rests entirely with the experimenter. The east door cannot be closed until the chamber lighting has been switched from 'main' (7 500 watts) to 'emergency' (90 watts). Anyone locked in the chamber can communicate with the control room by an intercommunication set, and can operate a fire-alarm-type switch which will stop the east door, restore the main lighting and open an interlock to prevent cyclotron operation.

(13) CYCLOTRON SUITE

The site at Hammersmith Hospital, where the M.R.C. machine has been built, is quite flat, and a careful analysis of costs indicated that it would be economic to house the cyclotron above ground. The building containing the cyclotron and associated plant (Fig. 16) is at one end a multi-storey block containing clinical facilities and research laboratories.

Radiation shielding for health protection requires between 4–6 ft of concrete surrounding the cyclotron.[8] A limit to the level of radiation from the cyclotron appearing in the main building was set as the equivalent of the radiation occurring naturally from cosmic rays, so that intermittent running of the cyclotron shall not unduly affect experiments using low-intensity radiation sources. To achieve this the radiation shield between the cyclotron and the main building has to be 10 ft thick, the remaining walls and ceiling being 6 ft thick.

A section of the 10 ft wall between the cyclotron chamber and an adjacent assembly bay can be moved sideways on roller races to give access to bulky equipment. The block weighs 190 tons and is rope-hauled by an electric winch. When in the closed position it fits into the surrounding concrete with a maximum clearance of $\frac{1}{2}$ in. When the block is moved aside it exposes a pit 4 ft deep, which is bridged by a two-leaf drawbridge. A rail track runs from the cyclotron chamber over the drawbridge to the assembly bay, and the whole electrode and vacuum assembly can be withdrawn on this track.

The personnel entrance to the cyclotron chamber consists of a passage through the 6 ft concrete wall which can be covered by an overlapping door consisting of a steel tank filled with steel-loaded concrete and weighing 35 tons. In the closed position the door uncovers a 12 in diameter water window through which the cyclotron may be viewed during operation.

(14) PERFORMANCE

This must necessarily be a brief review of the operation of the cyclotron since the machine is still being commissioned and it is intended to present a further paper on the operation at a later date.

At full radius 450 μA of both α-particles and deuterons have been circulated, with a $\frac{5}{16}$ in square beam-defining aperture[9] at the end of the first turn. An extraction efficiency of 20% has been measured at low level. When the final adjustments have been made to the size of the limiting aperture and to the deflector it is expected that an extracted beam current of 200 μA will be attained.

The principal techniques being used for aligning the beam rely on a measurement of temperature rise and flow rate in water-cooled copper targets and on the visual observation of the heating of carbon targets placed at various positions in the machine. The latter technique, which has proved invaluable, is made possible by the 10 observation ports on the sides of the dee-box. Through these ports any point in the path of the beam can be seen while the machine is running.

The deflector channel has not yet been accurately aligned. An attempt is being made to increase extraction efficiency by using beam-defining apertures to intercept the stray particles which would otherwise strike the dees, septum and deflector. Vertical restriction of the beam in this way has already been shown to reduce the temperature rise of the dees by about 20° C.

(15) ACKNOWLEDGMENTS

The authors wish to acknowledge their debt to members of the cyclotron team both past and present and to the many industrial firms who co-operated in the building of the present machine, particularly C. A. Parsons Ltd., who designed and built the magnet. A special debt of gratitude is due to their many colleagues in universities and elsewhere, both here and oversea, for their assistance with advice, drawings and information; and in particular, to those in the United States who so readily contributed many of the data upon which the present design is based.

They wish to record their thanks to the Director and members of the Radiotherapeutic Research Unit for their encouragement and collaboration throughout the project, and to the Medical Research Council for permission to publish the paper.

(16) REFERENCES

(1) GALLOP, J. W.: 'Notes on a Tour of American Fixed-Frequency Cyclotrons in the Autumn of 1950' (Medical Research Council Radiotherapeutic Research Unit, 1951).

(2) ADAMS, J. B., ROWE, G. W., and JAMES, D. B.: 'The Design of the Harwell Magnet', A.E.R.E. Report G/R 175.

(3) HOUGH, P. V. C.: 'Radial Oscillations in the Cyclotron', *Review of Scientific Instruments*, 1953, **24**, p. 42.

(4) ANDERSON, H. L., *et al.*: 'Synchrocyclotron for 450 MeV Protons', *ibid.*, 1952, **23**, p. 707.

(5) SCHMIDT, F. H., and JAKOBSON, M. J.: 'Cyclotron Oscillations and the Shifting Inter-Dee Ground Surface', *ibid.*, 1954, **25**, p. 136.

(6) LIVINGSTON, R. S., and JONES, R. J.: 'High-Intensity Ion Source for Cyclotrons', *ibid.*, p. 552.

(7) POST, R. J.: 'The Control System of the Medical Research Council Cyclotron'. In preparation.

(8) BOAG, J. W.: 'Radiation Protection for the Radiotherapeutic Research Building', Medical Research Council Radiotherapeutic Research Unit, Internal Memorandum, 1949.

(9) POWELL, W. B.: 'Improving the Characteristics of the Cyclotron Beam', *Nature*, 1956, **177**, p. 1045.

DISCUSSION BEFORE THE INSTITUTION, 21ST MARCH, 1957

Dr. Constance Wood: As a doctor whose work has been that of using radiations in the treatment of disease, I should like to take this opportunity of saying that I and my medical colleagues engaged in similar work feel a particular debt of gratitude to engineers, because it is you to whom we owe the design and construction of the apparatus which produces the radiations which we use in medicine.

Unfortunately, there is still little known of the cause of cancer, and in the present state of our knowledge at any rate, radiation plays a most important part in the treatment of malignant disease. It is, I think, true to say that in approximately two-thirds of all cases of malignant disease radiation, alone or in combination with surgery, is the treatment of choice.

It is the constant aim of the radiotherapist to improve the distribution of the radiation dose in cancer tissue, and inevitably, therefore, radiotherapists have been very interested in the produc-

tion of X-rays of higher and higher voltage. As you know, engineers have now produced apparatus which can be used in clinical medicine producing X-rays of many millions of volts. The linear accelerator, essentially a product of British engineers and physicists, is an example of such a machine.

We do not believe that the biological effect of super-voltage X-rays, however, differs very greatly from that of the conventional 250 kV X-rays, but it is well known that the biological effect of neutrons is more powerful than that of X-rays. With the beam of fast neutrons from the Medical Research Council's cyclotron we hope that it may be possible to make investigations which will help to determine the value of neutrons in the treatment of malignant disease. (There is no question of the efficiency of neutrons in destroying malignant cells; what is not known is whether this property can be turned to advantage as compared with other forms of radiotherapy in the treatment of malignant disease in patients.)

As the authors have said, this is the first cyclotron to be placed in a hospital, so facilitating the planning of any clinical investigation.

A second function of this cyclotron is to produce certain radioactive isotopes which cannot be produced in the atomic pile, e.g. iodine 124. It is well known that iodine 131, produced in the atomic pile, has for a number of years been used in the treatment of certain disorders of the thyroid gland. Iodine 124, which can be produced only in the cyclotron, when it decays emits X-rays, not β-rays, and the X-rays which it produces have a penetration appreciably greater than that of the β-rays emitted by iodine 131, so that the radiation dose to the tissue will be much more homogeneous. Thus the cyclotron-produced iodine 124 may, because of its different physical properties, prove to be more valuable in the treatment of thyroid disorders than the commonly used pile-produced iodine 131.

Again, many isotopes are of short life, and a great advantage of the M.R.C. cyclotron lies in its ability to produce in a hospital certain isotopes which, on account of their very short life, must be used immediately. Facilities have been provided adjacent to the machine for this purpose.

The third function of the cyclotron is to provide a source of radiation for radiochemical and radiobiological investigations in medical research.

I have tried to outline briefly the three main uses to which the machine will be put, and in closing I should like to pay a tribute to the M.R.C. cyclotron team, so ably headed by Mr. Gallop and Mr. Vonberg. They have had work to do of no little difficulty. They have been in a position very different from that which would have obtained had such a machine been constructed for use in a large industrial organization, with all the reserves that such an organization would have to fall back upon. I congratulate the team as a whole on their achievement of a job well done. I can only trust that the various uses to which the machine will be put will prove to be worthy of the efforts put into it by its designers and constructors.

Prof. J. Rotblat: I wish to point out that this is not only the first cyclotron in a medical institution but is a very considerable addition to the total number of cyclotrons in this country, since there are very few in this energy range. Up to now isotopes for medical use have been prepared in the other cyclotrons. This work will now be taken over, no doubt, by the Hammersmith machine, which I hope may sometimes be available for help in nuclear physics research.

As this machine will be used mainly for medical work, I should like the authors to say a little more about the facilities which have been provided for this sort of application. For example, we are told that it may be used for neutron therapy, and so I assume that it will be used to provide a beam of neutrons, but what sort of facilities have been provided to direct the beam of neutrons on the patient? Is it also intended to produce protons? Do the authors envisage any other direct use of the beam apart from the neutrons? What facilities are there to produce radioisotopes?

Mr. W. Walkinshaw: I should like to mention some recent research work which we have been doing on constant-frequency cyclotrons at Harwell. In the case of the M.R.C. cyclotron, the particle frequency decreases with increase in energy, so that there is a phase shift of the particle with respect to the radio-frequency, and this in turn limits the energy which can be obtained with a given dee voltage with machines of this type. As long ago as 1938 Thomas suggested a method of getting over this difficulty. Essentially his suggestion was to have a field which increases with radius; the magnetic field is divided into sectors, and in alternate sectors the field is made to increase and decrease above and below the main field. The result of this on particle motion is that instead of having a circular orbit we have a slightly distorted orbit, so that these particles cut the sector edge at a slight angle. This gives the well-known effect of edge focusing and restores the vertical stability.

This machine has a certain disadvantage, in that as the energy increases the flutter in the field becomes rather large. Nevertheless, in 1951 a cyclotron was designed on this principle at Berkeley, U.S.A., which accelerated electrons to about 0·4–0·6 times the velocity of light, an equivalent energy of 150 million volts for protons. It is, of course, possible to design a proton machine on these principles.

Two years ago Kerst, of betatron fame, suggested a further improvement in this design. He was investigating spiral-ridged fixed-field machines, but this can be applied to the cyclotron. Essentially it means that in the early Thomas design we increase the angle at which the particle goes to the high-field region, and we can therefore decrease the flutter appreciably.

We have been trying to apply this principle to the Harwell cyclotron but there are many difficulties and dangers. Unfortunately, as soon as we try to get something for nothing we come up against another difficulty and have to sacrifice something. We find that the tolerances become very much tighter than in existing machines, and in particular it is necessary to initiate the particles from the centre with very small oscillations round the mean values.

Although the design and construction of machines like the M.R.C. machine have gone on for 25 years, resulting in this very fine instrument, we still find ourselves ignorant of what is happening at the centre of cyclotrons. It is an extremely difficult problem. I know that the authors have contributed to this aspect, and I should very much like them to tell us something about the measurements which have been made near the centre of this cyclotron.

Mr. G. R. Newbery: I should like to mention a few of the measurements which Dr. Dyson, Mr. Wojcikiewicz and I have so far made with the radiation from this machine. The dee voltmeters (see Section 4.7) were calibrated by a method used by Greening* for X-ray machines, using the X-rays generated at the surface of the dees when struck by electrons arising from the dee-box by field emission. These electrons and the resulting X-rays have an energy up to the maximum of the dee-to-earth voltage, i.e. up to approximately 80 kV. By observing with two scintillation counters the ratio of the X-radiation transmitted to the X-radiation scattered by a tungsten foil, as a function of the dee voltmeter reading, it was possible to determine accurately the dee voltmeter reading at which the peak r.f. voltage corresponds to the K absorption edge of tungsten, i.e. 69 kV.

I should like to ask the authors to give an account of a pheno-

*GREENING, J. R.: *British Journal of Applied Physics*, 1955, **6**, p. 73.

menon which was observed during the dee voltage calibration, namely a very large low-frequency ripple on the r.f. voltage.

We have studied the fast neutron dose rate which results when a 100μA beam of deuterons strikes an internal beryllium target. A small homogeneous (CH/C_2H_2) gas-flow type of ionization chamber was used outside the cyclotron chamber, the neutron beam being collimated by a 6 in diameter hole through the 6 ft concrete wall. At 6 metres from the target and at 15° from the forward direction, the dose rate for 100μA of deuterons is approximately 1·5 rads/min in tissue. The γ-ray dose rate is only a few per cent of that due to fast neutrons. The fast neutron dose rate at 1 metre from the target in the forward direction is thus approximately 1 rad in tissue per microampere per minute. Measurements have so far been limited by the inability of the beryllium target to withstand more than 100μA of deuterons; perhaps the authors could explain this limitation and say what they are doing to overcome it.

During early measurements it was found that the cyclotron was causing interference with an adjacent television receiver, and it would be of interest if the authors could comment on this.

I should also like to ask the authors for an assessment of the difficulty, or perhaps I should say the ease, with which they can change the characteristics of the radiation produced by this machine, e.g. from a deuteron beam to an α-beam or to a proton beam, from low-beam current to high-beam current, or from an internal target to an external target. Finally, the authors state that this machine produces 15 MeV deuterons; is this in fact the maximum deuteron energy that can be achieved?

Mr. M. Snowden: I should like to ask the authors one or two questions about the magnetic field measurements which have been done on this cyclotron. First, with regard to the azimuthal uniformity, I notice that before adding shims the uniformity was 1 in 10^4 oersteds, but that after the shimming was done an adjustment was necessary. What was the final variation in azimuthal field and what was the dominant harmonic in this?

I notice in the paper that flaws were detected in the Armco steel, but that these did not produce any measurable effects on the magnetic field. However, the equilibrium plane was found to be as much as $\frac{3}{4}$ in high in the region of the central cone and $\frac{1}{2}$ in low at $12\frac{1}{2}$ in radius. Presumably this change was not due to mechanical tolerances. Is there not some other explanation for this?

In connection with the ion source, a current of 10 mA is mentioned. I should like to ask how this was measured, and how it varied with the dee voltage? Secondly, what was the optimum gas flow used, and was it in fact limited by the available pumping speed? Finally, it appears that the phase of the beam relative to the radio-frequency varies during the acceleration cycle. It would be rather fortuitous if the overall phase difference cancelled out. Is there not some reactive component due to beam loading which affects the r.f. tuning?

Mr. R. Bruce: Before the final choice of aluminium for the magnet windings two basic designs were compared, namely paper-insulated copper windings of disc-coil construction cooled by the circulation of transformer oil, and aluminium windings utilizing an extruded hollow-section conductor arranged for cooling-water circulation within the central hole. Both designs occupied the same total window area, but the copper windings had a relatively poor space factor due mainly to the space required radially for the flanged joints of the enclosing tank, coupled with the provision of oil cooling ducts. The gross aluminium section was 75% in excess of the gross copper section, which favoured the aluminium design on a power-loss basis by some 6%, after due allowance for the poorer conductivity of the aluminium.

The aluminium could be obtained in virtually any required length, being made by a continuous extrusion process not possible with copper. The lengths obtained were about 1 200 ft each, sufficient for a pair of disc coils, which were wound so that the crossover between discs at the inside was in the middle of the length of conductor. This method of winding eliminated all joints within the body of the winding.

Rather more than $3\frac{1}{2}$ tons of extruded conductor were used for each pole, and notwithstanding the disc coils being wound in pairs, the whole of the windings for one pole were dealt with without removal from the winding machine (see Fig. A). This was possible by the use of a prefabricated steel winding former attached to both the faceplate of the horizontal-axis winding machine and the winding endplate to which the windings were finally clamped. This former provided a rigid support to carry the windings during construction and to withstand the tension applied to the conductor while winding. On removal from the winding machine and after extracting the winding former, the self-contained windings for each pole were fitted with sheet aluminium covers on both inside and outside surfaces.

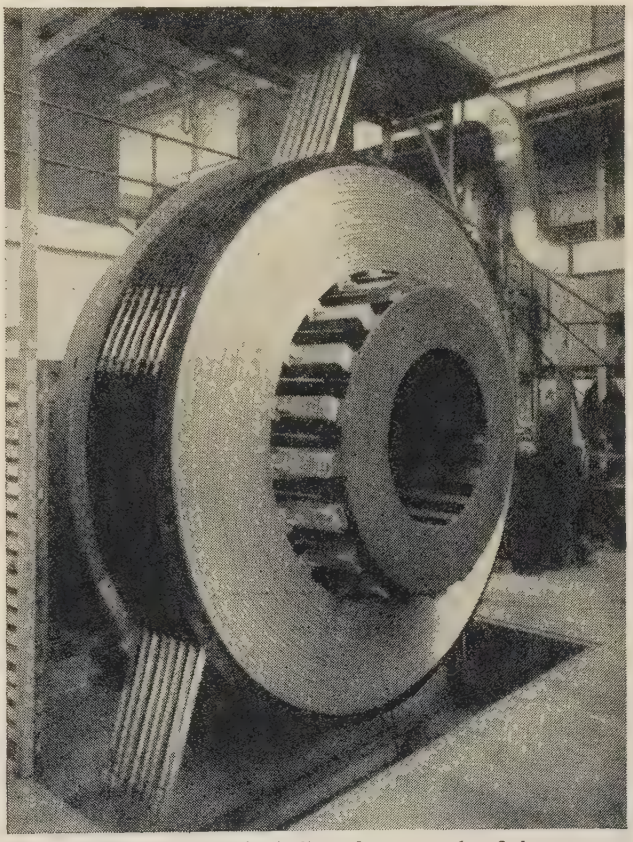


Fig. A.—Almost completed windings for one pole of the magnet on the winding machine.

Dr. J. W. Boag: We must all be impressed by the detailed design which has gone into every component of this machine. One elegant detail was the way in which the current was equally distributed between the various cooling tubes in the short-circuit by placing them along the lines of current flow in the homologous infinite-current sheet.

In Section 13 the authors state that the design of the 10 ft shielding wall between the cyclotron chamber and the main building was intended to cut down the cyclotron radiations to the level of cosmic-ray intensity in the main building. The specification was tighter than that; it called for a maximum increase of two counts per minute in a partially shielded counter in the main building. At least 9 of the 10 ft of concrete in the moving door were necessary to meet that specification.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. J. W. Gallop, D. D. Vonberg, R. J. Post, W. B. Powell, J. Sharp and **P. J. Waterton** (*in reply*): We should like to acknowledge, on behalf of the profession, the generous tribute paid to electrical engineers by Dr. Constance Wood, and to thank her for her explanation of the uses to which the cyclotron will be put for medical research.

Prof. Rotblat asks about the application of the neutron beam. Measurements by Mr. Newbery have shown that the neutron flux from an internal target is sufficient for the commencement of a trial of neutron therapy, with the patient outside the shielding at a distance of some six metres from the target. A hole in the shielding will provide a collimated horizontal beam. Neutron-induced activity in the cyclotron from the internal target is a serious handicap which will limit this use of the machine to about half an hour per day, or possibly less. A focusing system is being designed to carry an extracted beam to a target just inside the shielding, so reducing the neutron activation problem.

The machine produces 15 MeV deuterons, 30 MeV α-particles and 7·5 MeV protons from H_2^+ ions. It is expected that the external beam of α-particles will be of particular interest for radiochemistry and radiobiology. For radioactive-isotope production, static and rotating internal targets have been provided, and an external target, protected by a foil window, will be used for less stable materials. There is provision for pneumatic tube delivery of the target to the 'hot' workshop and radiochemistry laboratory. An initial batch of 30 millicuries of ^{22}Na has been produced in a continuous 80-hour bombardment, although priority will generally be given to short-lived isotopes for medical use.

In reply to Mr. Walkinshaw's question, we have used remotely controlled carbon stops and apertures viewed by telescope to locate and measure the first few turns of the beam for a number of ion-source and dee geometries. In each case there is good agreement between the experimentally determined and the calculated positions of the beam. Briefly, the calculation assumes a point of origin at the ion-source slit and an enhanced field between source and feeler, and allows for the reduction in electric field strength within the dee.

Further, it has been shown that the beam is partially coherent throughout the motion; for it is possible, by means of a stop placed at the first turn, to effect a considerable reduction in deflector heating without reducing the external beam current.

Mr. Newbery refers to an interesting phenomenon which puzzled us for some time. Owing to the deflection of the dee surfaces under the influence of electrostatic stress, the dee capacitance proved to be appreciably dependent on the dee voltage. A snap tuning effect resulted, which, in conjunction with the regulation of the h.t. supply, led to a mechanical oscillation of the dee surfaces and to a large low-frequency ripple on the r.f. voltage. The frequency of oscillation was mainly determined by the mechanical characteristics of the dees. The vibration was prevented by the use of a small prop (Fig. 9) near the centre of the dee mouth, placed so as to avoid the early turns of the circulating beam.

The limitation of 100 μA of deuterons on internal beryllium targets used so far is due to heat-dissipation difficulties in the beryllium. Stationary targets, $\frac{1}{32}$ and $\frac{1}{16}$ in thick, have a life of only a few minutes at 1 kW. Melting occurs at the surface and a crack forms, allowing water to leak into the vacuum vessel. Targets rotating at 100 r.p.m. show heavy cracking along the isothermal lines of the heat wave after a few hours' running at 1·5 kW. At 300 r.p.m. the cracking is much less evident under similar conditions.

Change of operation between 15 MeV deuterons, 30 MeV α-particles and 7·5 MeV protons simply involves change of ion-source gas and slight adjustment of magnet tuning. A change takes about half an hour to complete, owing to gas occlusion in the ion source resulting in initial contamination. Control of beam intensity will be by a remotely controlled iris at one complete orbit turn from the ion source.

By reshaping the magnetic field profile and providing modified dees, it is theoretically possible to increase the deuteron energy of the machine to 20 MeV, but such a change would result in much reduced beam current and loss of reliability.

In reply to Mr. Snowden's question on magnetic-field azimuthal uniformity, we can say that in the final field the first harmonic was dominant with an amplitude of about 1 gauss (slightly less at some radii).

The flaws in the Armco steel of the lids may have contributed to the variations in height of the equilibrium plane; but we think it more likely that it is due to mechanical asymmetry. A 0·001 in variation in flatness of a lid could give rise to a 0·75 in movement of the equilibrium plane and would only be responsible for about 2 oersteds change in absolute field strength.

The ion-source gas-flow for α-particles is 10–20 litre-microns/sec: higher flows result in de-focusing of the beam owing to change of ion-source conditions, and pumping speed is no limitation.

With a circulating beam of the order of 1 mA, presenting a radio-frequency conductance of about 5 micromhos to the dees, the susceptance was measured as $0 \pm 0{\cdot}25$ micromho. We are prepared to believe that this figure is fortuitous.

We are grateful to Mr. Bruce for the description of the construction of the magnet windings, which he designed. We are also grateful to another of our collaborators, Dr. Boag, for stating the exact requirements for the shielding, which he designed.

In conclusion, we would like to add that since the paper was presented the performance for 30 MeV α-particles has been brought up to 1·1 mA of internal beam, and 150 μA of external beam, at an extraction efficiency of 47%. Slightly higher deuteron currents have been obtained.

621.383

Paper No. 2388
Sept. 1957

PHOTO-ELECTRIC CELLS

A Review of Progress

By Professor J. D. McGEE, O.B.E., M.Sc., Ph.D., Member.

(1) INTRODUCTION

An understanding of the physics of the photo-electric effect[1] is necessary for the comprehension of the photoelectric cells and other devices which depend upon it. So the paper reviews briefly the experiments which led up to Einstein's theory and outlines the modern theory of the photo-electric effect. Then follows a description of the methods of making modern photo-electric cathodes, their characteristics, and those of a few of the electronic tubes in which they are incorporated. A brief account of solid-state theory introduces the internal photo-electric effect, and descriptions of the photo-conductive and photo-voltaic cells follow.

(2) HISTORICAL

The nineteenth century physicists established that when light was absorbed by matter, certain electrical effects were produced. In 1839 Becquerel[2] had found that, when light fell on two metal electrodes immersed in an electrolyte, a potential difference was established between them; this is now known as the photo-voltaic effect. Much later, in 1873, Smith[3] discovered the change in conductivity of selenium when sunlight fell upon it—the photo-conductive effect. Finally Hertz[4] discovered the photo-emissive effect—the loss of negative electric charges from metal surfaces which are illuminated. After the discovery of the electron by Thomson[5] in 1897, it was soon shown that it was electrons that were released by light from surfaces, and this explained why only negatively charged bodies could be discharged by the action of light.

The physical basis of the photo-voltaic and photo-conductive effects were quite unknown, and the main features of the photo-emissive effect could not be accounted for by the then current physical theories.

(3) EXTERNAL PHOTO-ELECTRIC EFFECT

The main characteristics of photo-electric emission are illustrated by reference to Figs. 1 and 2. In Fig. 1 an anode is electrically insulated from a photo-electric cathode and held at a potential relative to it by a suitable voltage source. If a beam of light falls on the cathode, electrons will be liberated from it and will, or will not, be collected on the anode, depending on the magnitude and direction of the electric field between it and the cathode, the current being measured by the meter. Under carefully controlled conditions the characteristics shown in Fig. 2 are obtained. Curve (*a*) shows the variation of anode current with anode potential when a fixed amount of white light, Φ, falls on the photocathode. This current is zero for a small negative potential OB, rises rapidly as the negative potential is reduced and becomes constant for a small positive collector potential, and remains constant as this potential is increased.

This shows that the photo-electrons have a small initial energy with a fairly precise maximum corresponding to the voltage represented by OB, since this is the potential required to prevent all of them from reaching the anode. Moreover, there is a definite maximum photo-current represented by the flat part of

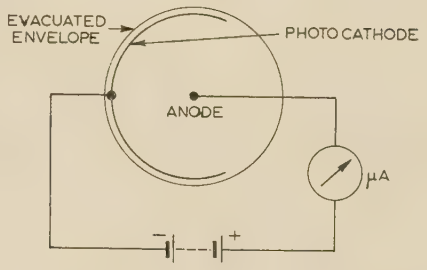


Fig. 1.—Simple photocell in circuit.

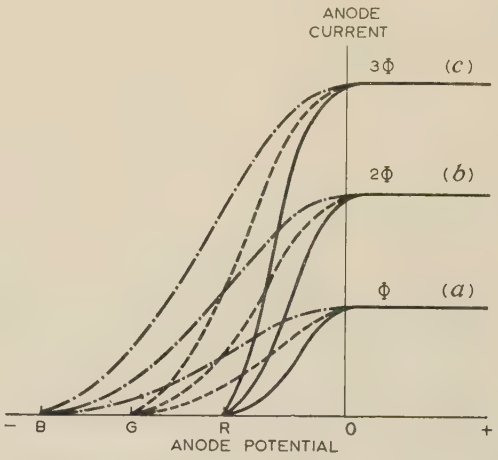


Fig. 2.—Idealized current/voltage characteristics of photocell.

curve (*a*), when substantially all the emitted photo-electrons are captured by the anode: this is termed the 'saturation current'.

Now, if the quality of the light remains the same, but its intensity is doubled curve (*b*) is obtained; it is similar to (*a*) in that it has the same cut-off potential OB, but it has twice the saturated current. Similarly, curve (*c*) shows the result for three times the amount of light. This leads to the surprising conclusion that increasing the intensity of the light increases the number, but not the initial energy, of the emitted photo-electrons. On the classical theories of interaction between light and electrons, an increase in energy of the electrons would have been expected.

If, now, only the blue component of the light is allowed to fall on the photo-cathode, and its intensity is adjusted to give the same saturation current at high positive anode potentials, corresponding to curve (*a*), the cut-off potential, OB, will be the same as before. However, if the green component is used it will be found that the cut-off potential, OG, is smaller. The same experiment repeated with the red component of the light gives a still smaller cut-off value, OR. These experiments show that the photo-electrons liberated have initial energies that decrease with increasing wavelength of the exciting light. This is the second fact that cannot be accounted for by classical theories.

Prof. McGee is Professor of Instrument Technology, Imperial College of Science and Technology, University of London.

The third difficulty arises from the fact that the photo-electric emission is almost instantaneous, even for very weak illumination, where it is clear that the light energy must be collected from quite a large area of the photocathode to provide enough to release one electron. It is difficult to imagine a mechanism to achieve this on the classical theory.

(3.1) Einstein's Quantum Theory

These difficulties were resolved in 1905 when Einstein applied[6] to the photo-electric effect the quantum theory of light,[7] by which Planck, in 1900, had accounted for the energy distribution of black-body radiation. This postulates that light in its interaction with matter acts as though it consists of indivisible bundles of energy, each of amount hf, where h is Planck's universal constant and f is the frequency of the light considered as a wave motion of wavelength λ and velocity c. In the photo-electric effect, such a bundle of light energy—a photon—is totally absorbed in the matter by an electron, which has its energy increased by an amount

$$V = \frac{hf}{e} = \frac{hc}{\lambda e} \text{ electron-volts} \quad . \quad . \quad . \quad . \quad (1)$$

If the universal values of h, c and e ($h = 6{\cdot}624 \times 10^{-27}$ erg-sec, $c = 3 \times 10^{10}$ cm/sec and $e = 4{\cdot}802 \times 10^{-10}$ e.s.u.) are substituted in this formula, and V is given in volts and λ in ångströms, we obtain the convenient relationship

$$V = \frac{12394}{\lambda} \simeq \frac{12400}{\lambda} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Hence photons of blue light (4000 Å) and of red (6000 Å) will impart energies of approximately 3 and 2 eV respectively to the electrons they excite.

It will readily be seen that Einstein's theory accounts completely for the anomalous effects noted above, since greater light intensity simply means more photons which are capable of exciting more photo-electrons, while photons of shorter wavelength represent bundles of greater energy which are capable of liberating more energetic photo-electrons. Since the energy arrives in bundles, the short time delay in imparting it to single electrons is also satisfactorily explained.

The photo-electric effect may occur in gases, liquids or solids, but in this review only the last of these will be considered. The light photons must be absorbed efficiently in a photo-electric material, and hence it can be neither transparent nor highly reflecting. Within the small distance that the light penetrates into the solid, it produces the excited photo-electrons which may produce internal photo-electric effects—photo-conductive or photo-voltaic—or they may escape from the surface to give the external photo-electric effect.

While the quantum theory gives the energy imparted to an electron by a photon within a solid body, it still remains for the electron to escape from the surface.

The free, or conduction electrons inside a solid body may be considered to have a distribution of energies, W_e, ranging from zero to a fairly precise maximum, $\hat{W}_e$, as shown in Fig. 3. Although, as stated, this is a fairly sharp cut-off at ordinary temperatures for metal surfaces, for some surfaces at room temperature, and for all surfaces at elevated temperatures, a proportion of the electrons have sufficient energy ($> W_a$) to surmount the potential barrier and escape from the surface. This is termed thermionic emission, and because of their inherently low work functions, it is usually appreciable from photo-electric surfaces and increases as the work function decreases.

If one of the free electrons of a solid absorbs the energy of a photon, the maximum energy that it can have in consequence is

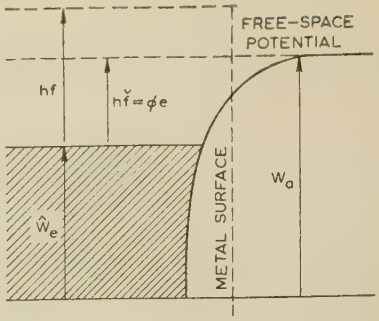


Fig. 3.—Energy relationships at surface of solid.

$\hat{W}_e + hf$; and if this is greater than the energy barrier at the solid surface, W_a, and the electron happens to be moving towards the surface, it will be able to escape with energy $\hat{W}_e + hf - W_a$. To obtain photo-electric emission it is necessary that

$$\left.\begin{aligned} hf &> W_a - \hat{W}_e \\ hf &> e\phi \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where ϕ is the work function of the surface. Thus, no photo-electrons can escape from the surface for light of frequency less than $\check{f}$, or wavelength greater than $\hat{\lambda}$, where

$$\hat{\lambda} = c/\check{f} = ch/e\phi \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This is known as the long-wave threshold, and is an important characteristic of photoelectric surfaces. Furthermore, the maximum energy of escaping photo-electrons due to light of frequency f is given by

$$V = \frac{hf}{e} - \phi \quad \text{electron-volts} \quad . \quad . \quad . \quad (5)$$

The additional velocity imparted to an electron by a photon will be random in direction, and hence only a minority can have a component of velocity normal to the surface sufficient to allow them to escape. Again, many will lose energy in collisions before reaching the surface and hence may not escape, or will escape with reduced energy. Thus we have the further conditions for the efficient liberation of photo-electrons that the light should be absorbed quite near to the surface and that there should be few free electrons with which the excited electron can make collisions which result in loss of energy.

(3.2) Practical Physical Considerations

The physical conditions required of a good photo-electric material can now be listed, and it will be seen that they restrict the possible materials very considerably: they are

(*a*) It can be neither a good reflector nor transparent, but it must allow light to penetrate to a depth of a few hundred atomic diameters.

(*b*) It should be neither a good insulator nor a good conductor, but it must conduct well enough to allow the small photo-electric emission current to flow.

(*c*) It should have few free, or conduction, electrons, so that excited electrons have a long free path.

(*d*) It must have a low-work-function surface.

(*e*) The atoms of the material should have a low ionization potential.

From these considerations it appears that semi-conductors are likely to be the most suitable materials. However, the earliest photo-emitters were common metal surfaces, and these must be considered briefly.

(3.3) Pure Metal Surfaces

The fact that the early experimenters had to use ultra-violet light to obtain photo-electric emission from the clean surfaces of common metals is now known to have been due to the high work function, 3–6 volts, of these surfaces.

It is clear from eqns. (2) and (4) that light of wavelength greater than 4000Å cannot release photo-electrons from a surface with a work-function greater than 3·1 volts. It was soon found that the alkali metal surfaces were more efficient photo-emitters than the common metals, especially for light of visible wavelengths, and this was later explained by the developments of atomic physics which showed that these elements have low ionization potentials in the gaseous state, and low-work-function surfaces in the solid state. The values for these characteristics of the alkalis and a few other common metals are given in Table 1.

Table 1

Element	Li	Na	K	Rb	Cs	Mg	Ca	Ni	W	Pt
Ionization potential (gas), volts	5·4	5·1	4·3	4·2	3·9	7·6	6·1	7·6	—	—
Work function of solid surface, volts	2·3	2·5	2·3	2·1	1·9	3·6	2·7	5·0	4·58	6·3

It is clear that both the ionization potential and the work function of the alkali metals decrease as the atomic weight increases, and that they are both low by comparison with other elements. Without discussing atomic structure in detail, it may be mentioned that this is because the alkali metals, belonging to the first column of the periodic table, have a single electron in the outermost orbit. This is relatively easily detached, and the more so the heavier the atom, thus giving the characteristics of low ionization potential and great chemical activity.

From these data it could be deduced that caesium would be the most efficient photo-electric metal, especially in the visible light range, and this in fact has proved to be so although it is always used in combination with other elements. In fact, all technically important photo-electric surfaces are so complex that no tidy theoretical explanation of their operation can be given. All that can be done is to indicate the possible function of their several components.

(3.4) Adsorbed Alkali Layers on Metal Surfaces

Clean metal surfaces are not efficient photo-electric emitters, and the process of forming a photo-electric cathode usually amounts to forming a thin composite film on a conducting metal substrate. A qualitative explanation of this enhanced photo-emission is given by the experiments of Taylor and Langmuir,[8] and Becker,[9] who showed that

(*a*) If caesium atoms are deposited from vapour on a clean tungsten wire, they will leave when vaporized off by heating, as positive ions. This happens only when about one-eighth of the surface is covered with caesium. For greater coverage with caesium a larger proportion come off as neutral atoms.

(*b*) The absorption of a monatomic layer of caesium on a clean tungsten surface decreases its work function from 4·5 to 1·54 eV. As more layers of caesium are deposited, the work function increases again to that of pure caesium metal (1·9 eV).

The interpretation of these observations is that the tungsten surface with its high work function (6·8 eV), and therefore high affinity for electrons, attracts into the surface the loosely bound valence electrons of the caesium atoms, leaving the resultant ions tightly bound to the surface by the electrostatic forces between their positive charges and the negative electron cloud in the metal surface. A partial layer of such ions on a surface results in an electric double layer with the positive charge on the outside, which tends to facilitate the escape of electrons from the surface, i.e. to reduce its work function. After a certain density of caesium atoms has been reached, the work function will have been reduced to such an extent that further caesium atoms are not ionized, but are only polarized, with the valence electron attracted towards the surface. These atoms will further decrease the work function of the surface, but the effect of additional layers will decrease until, for a thickness of perhaps 10–20 atomic layers, the work function will have risen to that of pure caesium.

This model satisfactorily accounts for the observed increase, followed by a decrease, of the photo-electric threshold wavelength when very thin layers of alkali metals form on a metal surface.

It appears convenient to distinguish between two components of the photo-electric emission: that stretching from the threshold wavelength through to the short ultra-violet being termed the 'normal photo-electric effect', and the fairly sharply defined peaks in the curve which are obviously due to especially efficient emission of photo-electrons due to light of a definite wavelength, which is known as the 'selective photo-electric effect'.

More detailed experiments have shown that

(*a*) The normal photo-electric effect is the result of the absorption of light in any state of polarization in the base metal. The escape of these electrons is facilitated by the lowering of the work function of the surface by the adsorbed film, and it is this which determines the threshold wavelength of the photo-electric emission.

(*b*) The selective photo-electric emission, resulting in peaks in the equal-energy response curves, is due to photo-ionization of the adsorbed atoms of the surface layer by light polarized parallel to the plane of incidence, and incident obliquely on the surface, i.e. it has a component of the electric vector normal to the surface. The underlying metal affects the position of the selective maximum only by the extent to which it modifies the binding forces of the peripheral electrons of the adsorbed atoms.

(3.5) The Effect of Oxygen adsorbed on a Surface

As will be seen later, a layer of oxygen is used in many photo-electric surfaces and, in general, it is difficult to prevent a layer of oxygen from forming on metal surfaces, even in a good vacuum. It is also known that such a layer of oxygen increases the work function of a surface and hence it would seem to be most undesirable on a photo-electric surface.

An explanation of this seeming paradox is given in an experiment by Kingdon,[10] who found that, if a piece of clean tungsten wire is oxidized by heating in oxygen to about 1 900° K, the work function is increased. Exposure of the oxidized wire in caesium vapour, so that a layer of caesium is formed on the surface, has little effect until the wire is flashed at about 1 600° K, when the work function falls to 0·71 volt—even lower than that of the clean tungsten surface with a caesium layer (1·54 volts). It appears that at the high temperature the caesium and oxygen atoms can rearrange themselves in such a way as to produce at the surface a very strong dipole field that reduces the high work function of the tungsten surface. This lowering of the work function shifts the threshold wavelength well into the infra-red and provides one necessary condition for efficient photo-emission. Similar experiments with silver as the base metal and with a lower-temperature thermal treatment gave a similar

shift of the threshold wavelength, but also a strong selective maximum in the blue.

(4) MEASUREMENT OF PHOTOCELL CHARACTERISTICS

The one characteristic that is required for every cell is its current sensitivity to light, expressed either in amperes per lumen or amperes per watt of light power. Other characteristics, such as the response in current to equal amounts of light power of different wavelengths, the current amplitude response to light input varying in intensity at different frequencies, the current output related to light input, etc., are fairly constant between cells of the same type, and except where a high degree of accuracy is required, the typical characteristic is taken to hold for all cells.

Since the photo-electric response varies a great deal with the wavelength of the light used, it is clearly important to know the composition of the light used for measurement. Such a standard source of light must also, if possible, be readily available, easily maintained, stable and long-lived. Such a light source is a tungsten-filament lamp run with a filament temperature of 2875° K (2870° K in the United States). A lamp of about 100-watt rating can be calibrated to operate at this filament temperature for given filament ratings, and its light radiation under these conditions can be measured; its energy distribution with wavelength is given by curve (*e*), Fig. 5. A light source of x candle-power radiates $4\pi x$ lumens isotropically, and hence the light, Φ lumens, accepted by a circular aperture of diameter y cm at a distance Z cm is given by

$$\Phi = \frac{\pi}{4} x \frac{y^2}{Z^2} \text{ (provided that } Z \gg y) \quad . \quad . \quad . \quad (6)$$

It is usually convenient to mount the lamp in one compartment of a box which is well ventilated, matt black internally and provided with screens to prevent reflected or scattered light from reaching the limiting aperture, which is mounted at a known distance from the lamp filament in a partition dividing the box into two compartments. The photocell whose current is to be measured is mounted in the second compartment, which is completely light-tight, except for the aperture admitting light from the standard lamp. The light passing through the aperture, given by eqn. (6), must fall on the photo-cathode of the cell, which is preferably positioned so that the light covers as much of the cathode area as possible. The size of the aperture and its position relative to the lamp can be adjusted to suit cells of various types and sensitivities. The cell must be operated under current-saturation conditions if it is a vacuum cell, and under standardized conditions for other cells.

The measurement of the equal-energy response curves, such as those shown in Fig. 5, requires much more elaborate apparatus and is usually left to standards laboratories, such as the National Physical Laboratory. A monochromator is used to give a beam of light, of a quite small range of wavelength, which is projected first on to the photo-cathode under measurement and the resulting photo-electric current is measured, and then on to a bolometer which measures the total light power in the beam. From such measurements the relative response to equal amounts of light power of different wavelengths can be plotted.

A useful method for obtaining a rapid check on the general shape of a colour-response characteristic can be provided fairly simply by using a series of narrow-pass-band light filters, such as interference filters, which can be calibrated for the amount of power passed from an agreed light source. This is a simple and rapid method for obtaining, say, ten points on the characteristic curve between, say, 3500 and 8000 Å.

There is no single method for the measurement of frequency response, although, in general, it consists in modulating the incident light at a known rate and amplitude and observing the corresponding variation of output current with frequency. Since the frequencies of light fluctuations requiring measurement, and to which cells will respond, vary from a few cycles per second to 10^3 Mc/s or more, it is evident that a great variety of methods must be used and a description of them is outside the scope of the paper. Similarly, other special characteristics such as fatigue, lack of linearity, temperature dependence, etc., will not be discussed.

(5) PHOTO-ELECTRIC CATHODES

The embodiment of these physical principles in practical photo-electric cathodes, the technique of their formation and the operational characteristics of the resulting photocells will be described in the Sections which follow.

(5.1) Pure Metal Surfaces

These must be kept very clean and free from oxide, and are often scraped clean before use. Even so, the work functions of such surfaces are high and the photo-electric threshold is usually in the ultra-violet region, ranging from cerium with a threshold at 4300 Å to platinum at 2000 Å. Thus the response to visible light is almost nil, and it is not high to ultra-violet. They are of little practical importance now, except where it is desired to have a photo-cathode which is insensitive to visible light but sensitive to ultra-violet light, or where it is convenient to avoid the use of the alkali metals because of their extreme chemical activity.

(5.2) Alkali-Metal Cathodes

These are usually made by evaporation of the bulk alkali metal *in vacuo*, so as to obtain a very clean surface. Sodium metal can be introduced into a lime-soda glass envelope by electrolysis through the glass wall. An electron-emitting hot filament in the evacuated bulb keeps the internal walls at a negative potential, and the bulb is immersed in a bath of sodium nitrate which forms the anode. The positive sodium ions pass through the glass and collect on the inner walls of the bulb.

These cathodes are the most efficient of those made from pure metals and their thresholds extend well into the visible spectrum. Their overall response to visible light is low, although over short ranges of wavelength they can have high efficiency. They are of very little practical importance.

(5.3) Alkali Hydride Cathodes

The surfaces of alkali metals can be activated by passing a discharge in a low-pressure atmosphere of hydrogen introduced into the bulb, the alkali metal being held negative and the anode positive. The hydrogen ions produce a hydride on the surface of the alkali metal, which turns blue in colour, and the sputtering or heating due to the discharge causes a thin layer of alkali atoms to be deposited over this hydride. The most useful of these is the potassium-hydride surface.

This surface is at least an order of magnitude more efficient than the clean alkali-metal surface. It has a maximum response in the blue and ultra-violet part of the spectrum and a very low dark current. On this account it was of importance for measurements in this wavelength range, but it has now been superseded by the intermetallic photo-cathodes.

(5.4) The Ag–O–Cs Cathode

The photo-electric surface developed by Koller[11] in 1929, consisting of a complex mixture of silver, oxygen and caesium, was the first of the really efficient modern photo-cathodes and it is still in use, especially when red or infra-red sensitivity is required. It can be made as an opaque, or as a transparent, layer, and although the former is the more efficient by perhaps

50%, the latter is so convenient in many practical applications that it is now probably more generally used. It is sufficiently important and typical to warrant a fairly full description of its manufacture, first as an opaque, and secondly as a transparent, cathode.

The size and shape of such a photo-cell may be very varied, depending on the particular application for which it is intended, and two representative designs are shown in Fig. 4, that shown in Fig. 4(*a*) normally having an opaque photo-cathode and that

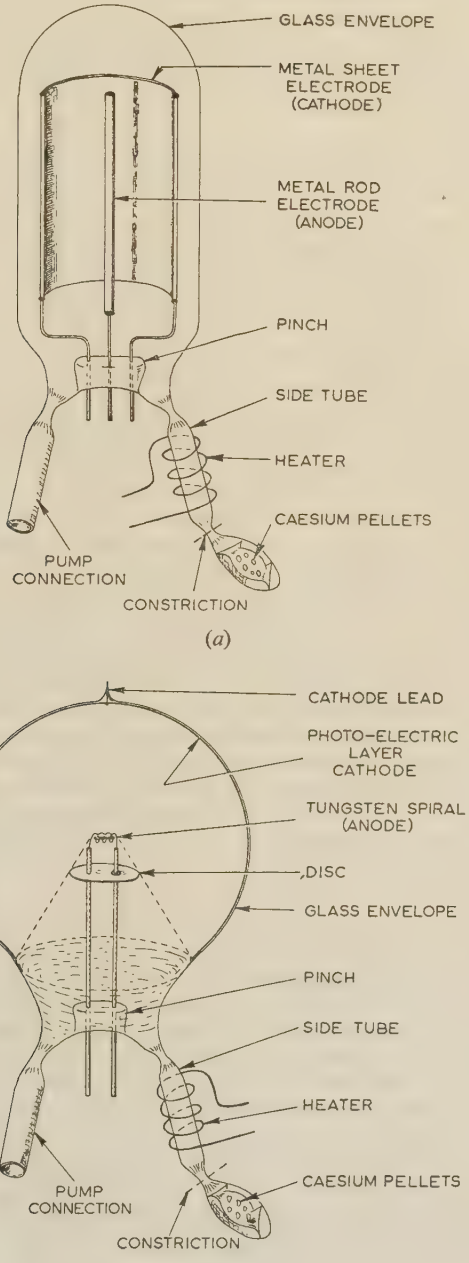


Fig. 4.—Processing of vacuum cells.
(*a*) Opaque cathode.
(*b*) Transparent cathode.

in Fig. 4(*b*) a transparent cathode; it will be as representatives of these two important types that the processes of making them will be described. In each case they are shown with the appendices attached that are required during processing.

The cell shown in Fig. 4(*a*) comprises a cylindrical glass envelope to which is sealed a pinch carrying three wires. The centre of these carries a metal rod about 1 mm in diameter, and the other two support a semi-cylindrical metal sheet; this may be of solid silver, but it is usually of cheaper metal, such as a stainless steel or nickel, with a thin layer of silver electroplated or vacuum-evaporated on to the surface facing the central electrode. The envelope is evacuated through the left-hand tube, and in the other tube is a capsule containing several pellets from which metallic caesium can be generated. These pellets are pressed from an intimate mixture of caesium chromate with very finely divided aluminium (or silicon or zirconium) and a proportion of fine tungsten (or tantalum) powder. On being heated to about 500° C, the caesium chromate and the aluminium powder react exothermically, liberating free metallic caesium. The reaction would be explosive but for the inert tungsten powder, which acts as a moderator and enables the reaction to be controlled by adjusting the externally applied heating: this is usually performed inductively by an eddy-current heater. As the metallic caesium is formed it diffuses out of its capsule; it is liquid at room temperature, with a high vapour pressure at 200–300° C, and can be distilled past the constriction and condensed again in the portion within the heater coil. The unwanted caesium-generating capsule can then be removed by sealing off the glass tube at the constriction, but the metallic caesium is not generated until other processes have been carried out. It is very important that all metal and glass parts should be extremely clean, and the metal parts should, where possible, be vacuum-stoved at a high temperature before use. The materials used, such as silver and the ingredients of the caesium-generating pellets, must be of the highest possible chemical purity.

Before the metallic caesium is generated, the whole tube should be evacuated and baked for an hour at 400° C, and all metal parts should be eddy-current heated where possible. Good vacuum conditions are also essential for success.

A small pressure of 0·1–1·0 mm Hg of oxygen is admitted to the tube through the pump connection by heating a bulb containing a suitable oxygen-producing material such as potassium perchlorate, barium peroxide, etc., the exact pressure of oxygen depending on the size and geometry of the tube. The silver surface of the sheet electrode is then oxidized by a discharge in the oxygen: this may be produced by a direct voltage applied between the electrodes with the sheet electrode negative, so that positive oxygen ions are attracted to it and react with the silver surface to form silver oxide; alternatively a glow discharge is produced in the oxygen by bringing close to the tube an external electrode energized by a high-voltage h.f. supply. The latter method is now preferred, because the distribution of the discharge can be adjusted to some extent by positioning the external electrode, and hence a more uniform layer of oxide can be laid down on the silver surface; this helps in achieving a more uniform photo-electric surface.

Silver is not oxidized by heating in air or oxygen; in fact, silver oxide is reduced when heated to about 200° C. As the oxidation progresses, the metallic surface of the silver passes through a more or less repeated cycle of colours: straw, red, blue, green, yellow, blue, blue-green. The oxidation is usually stopped at this stage and the remaining oxygen is pumped away. The tube is then baked at 200° C for a few minutes to remove excess adsorbed oxygen, and the metal parts are flashed and the glass seal-off constrictions are heated to softening point for the same purpose. The metallic caesium is then released as described above.

The tube is then covered with a baking oven which has a glass or mica window through which a constant beam of light

can be projected on to the concave surface of the sheet electrode, and a p.d. of 100–200 volts is applied between the electrodes, the central rod being positive. A galvanometer (and protective resistor) are included in the circuit.

The whole assembly is then heated to about 170° C, at which temperature the heating coil surrounding the side tube containing the caesium is switched on; this tube is thus heated a further 50° C, and caesium vapour begins to diffuse slowly into the envelope, where it reacts with the silver oxide to form caesium oxide and free silver, thus

$$Ag_2O + 2Cs = Cs_2O + 2Ag$$

At this stage the galvanometer will begin to pass a current which may be due to one or all of three effects, namely

(*a*) Genuine photo-electric emission: this can be checked by switching the test lamp on and off.

(*b*) Thermionic emission from the photo-cathode: this is high at the baking temperature, but is independent of the illumination and the applied voltage, provided that the emission is saturated; hence, any residual saturated current after the light is switched off is thermal emission.

(*c*) Electrical leakage current: this often arises from the condensation of caesium vapour on the insulating surfaces between the cathode and the anode; it obeys Ohm's law, and the proportion of the total current due to it can be estimated by, say, doubling the cathode-anode p.d. and noting the increase in current.

By making successive observations with and without illumination on the photo-cathode, and with normal and double the saturating voltage, a skilled operator can follow the growth of the three sources of anode current. Usually the electrical leakage current will be large at first and the other two small; then, as the caesium is absorbed by the silver oxide, the photo-emission and thermionic currents grow while the leakage current decreases. The process is continued until the photo-electric current has passed through a maximum and begins to fall. The caesium side-tube heater is then switched off, but the baking of the envelope may be continued for some minutes to complete the absorption of the caesium by the silver oxide. A very delicate balance between the silver oxide and caesium must be achieved to obtain a highly efficient photo-emissive surface. The colour of the surface under known conditions is one guide to the progress of the activation, and it is also found that if the sensitivity increases on cooling the cell has enough caesium, while a decrease of sensitivity indicates insufficient caesium. Excess caesium may be removed by continued baking to distil it slowly out of the envelope, or it may be absorbed on a layer of material that has a strong affinity for caesium, such as Aquadag or tin oxide, and is placed in a position in the tube not readily accessible to caesium vapour. However, in this type of cathode, serious overdosage with caesium can seldom be corrected satisfactorily.

It should be noted that in commercial photocells, in which the conditions of manufacture can be very strictly controlled, the caesium capsule may be located in the glass envelope and the caesium generated within the tube. By careful adjustment of the amount of oxide and the amount of caesium generated, adequate sensitivity can be achieved.

At this stage the processing of this type of cell is complete. The caesium side-tube is sealed off and the cell is sealed off from the pumps.

(5.4.1) Transparent Ag–O–Cs Photo-Cathode.

It is frequently a great convenience to admit the light to the photo-electric layer through a transparent support on which the layer is formed, and to remove the emitted electrons from the free surface of the layer. Such a layer is described as a transparent photo-cathode, and its preparation for silver-oxygen and caesium will be described with reference to Fig. 4(*b*).

Several features are common with that shown in Fig. 4(*a*), but in this example a spherical glass bulb is used and the pinch carries only two wires. These support a spiral of tungsten wire, which is wrapped round a short cylinder of silver and can be heated electrically to melt the silver and evaporate it on to the walls of the glass envelope. One of the wires carries a disc, so placed that it casts a shadow in the silver deposit on the glass wall in the region of the pinch and so preserves insulation between the anode and the cathode. The cathode lead is a wire sealed through the wall of the envelope approximately opposite the pinch, and it is attached to the inner wall of the envelope in such a way that it makes good contact with the silver layer when it is evaporated from the bead. Provision is again made for a pumping stem and a side tube in which caesium is generated from a capsule and warmed by the local heating coil.

After evacuation, baking and degassing of the tube, the silver is evaporated from the bead on to the glass wall to form the base layer for the photo-cathode. This could be an opaque layer, in which case the procedure is as described above. However, to make a transparent photo-cathode a layer of silver is formed that absorbs about 50% of the transmitted light. Oxygen is then admitted and a discharge is excited in the gas until the blue-grey silver layer has become quite colourless. No colour changes are visible in this case. Activation with caesium vapour is now carried out as for the opaque surface, and the blue colour of the original silver returns to the photo-cathode as the caesium combines with the silver oxide and frees metallic silver. Also the activated layer becomes a reasonably good conductor, as was the original silver layer, while the oxidized layer was highly insulating.

The original process as developed by Koller stopped at this point and the complete tube was sealed off. However, Asao and Suzuki[12] discovered that, if a further layer of silver is evaporated from the silver bead on to the activated surface, the overall sensitivity to white light rises by a factor of 2–4. At the same time the threshold wavelength decreases to about 8000 Å from about 12000 Å and the dark current also decreases. However, this change is not permanent; the threshold wavelength and the dark current slowly return to their normal values, but, fortunately, the enhanced photo-electric efficiency is permanent. The effect of the additional silver may be completed by baking the tube to 170° C for a few minutes. This additional silver layer may be applied to an opaque Ag–O–Cs cathode with similar beneficial results.

The characteristics of these opaque and transparent cathodes are very similar, except that the former is about 50% more efficient than the latter; the former may reach sensitivities of 60 μA/lumen to the light of a standard tungsten-filament lamp, while the latter reaches only 40 μA/lumen. The long-wavelength threshold is at about 12000 Å, and there are maxima in the equal-energy response curve [Fig. 5(*d*)] at about 8000 Å and in the ultra-violet region. It will be noticed that the efficiency is at its lowest in the visible region of the spectrum. Thus, it is the most sensitive photo-electric cathode for use in the near infra-red region, of wavelengths between 7000 and 12000 Å. It follows from the long-wavelength threshold that the work function of this surface is very low (< 1 volt), and hence, as is to be expected, the thermal emission is large at normal temperature, being about 10^{-10} amp/cm^2. This is very dependent on temperature and is greatly reduced by cooling the surface, the reduction being by a factor of about 10^3 at −80° C.

(5.4.2) Very Thin Ag–O–Cs Photo-Cathode.

If the basic layer of silver is made so thin that it does not become a conductor, and if this is activated in the usual way, as described above, it is found to have a fairly good photo-sensitivity

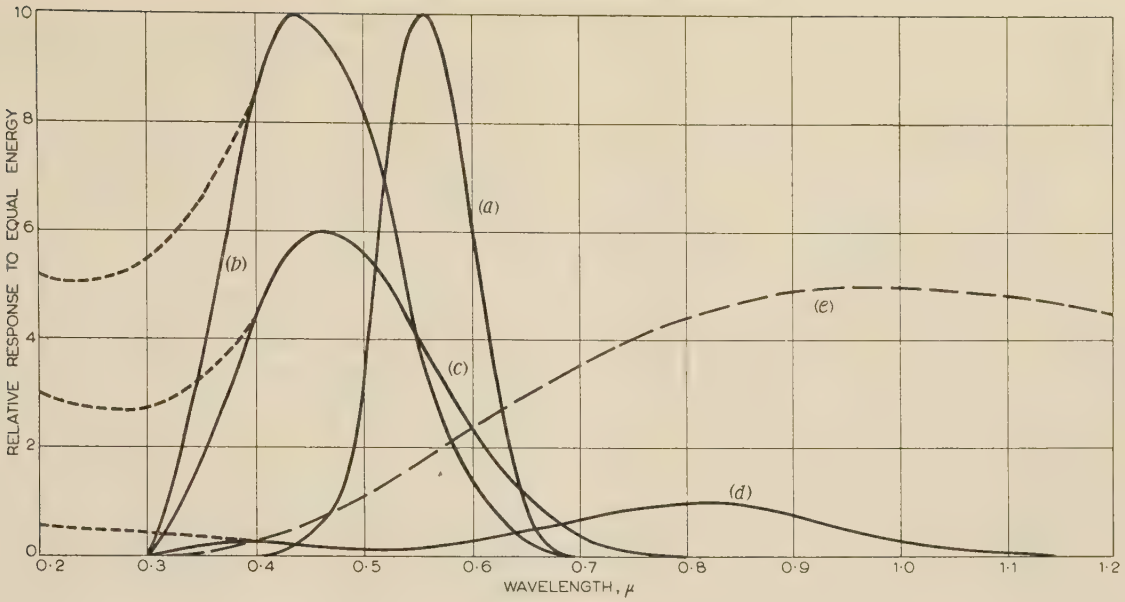


Fig. 5.—Equal-energy colour response curves.
(*a*) Human eye. (*b*) Sb–Cs cathode. (*c*) Bi–Ag–Cs cathode. (*d*) Ag–O–Cs cathode. (*e*) Energy emission.

and an equal-energy response curve which is almost wholly restricted to the visible region of the spectrum. This approaches more nearly the response of the human eye and was therefore of great importance for television camera tubes before the advent of more modern surfaces. It was used in the early Emitron television pick-up tubes for the photo-sensitive mosaic surface.[13]

(5.5) Intermetallic Compound Surfaces

These were discovered by Goerlich[14] in 1936 when searching for an efficient transparent cathode—at that time very urgently required for television pick-up tubes. He found that the alkali metals formed stable alloys with certain metals, notably antimony, bismuth, arsenic, thallium, gold and lead, to give photo-sensitive surfaces; of these, antimony in combination with caesium proved to be the most efficient. While these substances were at first thought to be alloys, there is now considerable evidence that they are, in fact, compounds of the two metals combined in stoichiometric proportions to form intermetallic compounds, such as $SbCs_3$.

(5.5.1) The Sb–Cs Cathode.

The first of these to be considered, the antimony-caesium cathode (henceforth Sb–Cs) may be made opaque or transparent by processes very similar to those employed in making the Ag–O–Cs surfaces, except that the oxidizing of the base layer of antimony is omitted. If an opaque cathode is required, a thick layer of antimony is evaporated (after the tube has been thoroughly evacuated and baked) from a bead of antimony or antimony alloy on to a metal supporting plate or the glass wall of the tube. If a transparent cathode is required, the thickness must be controlled fairly accurately to absorb about 40% of transmitted light, since the finished layer must be thick enough to absorb a fair proportion of the incident light but thin enough for the excited electrons to emerge from the free surface. The thickness may be controlled either by measuring the light absorption of the antimony layer as it is evaporated or by complete evaporation of an antimony bead of carefully controlled weight.

After the formation of the antimony substrate, the tube is baked and caesium is admitted to the tube to react with the antimony in the manner described for making the Ag–O–Cs cell. In this case the amount of caesium is not quite so critical, and if excess is admitted it can be compensated for by prolonged baking, which may remove the excess from the tube or cause it to be absorbed in a layer of material specially provided for this purpose, as was described in Section 5.4.1. It is probable that the ideal state for the surface at this stage is that it should consist of a layer of the compound $SbCs_3$ covered with a layer of caesium a few atoms thick.

It is often the practice to admit a little oxygen to the tube at this stage; this undoubtedly oxidizes some of the free caesium, and by some mechanism that is not understood, but is possibly a reduction of the work function of the surface, the sensitivity can be increased by up to five times. Great care must be exercised in admitting the oxygen, since excess can reduce, or can completely destroy, the sensitivity. The increase in sensitivity is usually accompanied by an increase in the threshold wavelength, which indicates a reduction in the work function of the surface.

On completion, the cathode should have a reddish colour by transmitted light, and a clear ruby colour is usually regarded as the optimum. Thus, the layer is fairly transparent to red light, which is in accord with the fact that its sensitivity to red light is very low, while it is opaque to blue light, for which it has the maximum photo-sensitivity. The equal-energy colour response curve for a transparent Sb–Cs surface is shown as curve (*b*) in Fig. 5; there is a pronounced maximum between 4000 and 4500Å and rapid falls to zero at about 6500 and 3000Å. While the fall in sensitivity towards the longer wavelengths is a real characteristic of the surface, that towards the short wavelengths is due mainly to the absorption of light by the glass envelope. If a quartz window, transparent to ultra-violet light is provided, this photo-cathode has a good response down to 2000Å, as shown by the dotted extension of curve (*b*). The opaque cathode is generally found to have its peak response further to the violet end of the spectrum than the transparent layer. This might be expected from the consideration that blue light is more strongly absorbed than red, and hence nearer the point of entry into a surface. Thus the electrons excited by blue light have a better chance of escaping from the surface of an opaque cathode through which the light enters than from that surface of a transparent cathode from which the light emerges.

The sensitivity of this photo-cathode may be 100 μA/lumen for an opaque, and 60 μA/lumen for a transparent, surface and

the threshold wavelength is about 6 500 Å, with the consequence that the dark current is very low compared with that of the Ag–O–Cs surface, being less than 10^{-15} amp/cm^2 at 20° C.

At the wavelength of maximum efficiency (~4 500 Å) the opaque Sb–Cs cathode can reach a quantum efficiency of 25%, and even the transparent cathode can often be between 10 and 15% efficient.

It may appear strange after what has been said earlier that an 'alloy' of two highly conducting metals should be a good photo-cathode. This anomaly is resolved by the realization that the 'alloy' is really a compound, $SbCs_3$, which has 10^6 times the resistivity of antimony and 10^7 times that of caesium. Thus the compound appears to be an intrinsic semi-conductor, and the mechanism of its photo-sensitivity appears to be due to a highly efficient internal photo-electric effect and an easy escape of photo-electrons through the surface.

While the Ag–O–Cs surface has been shown to exhibit metallic conduction characteristics, i.e. decrease in conductivity with increasing temperature, the $SbCs_3$ layers behave in the reverse manner, i.e as semi-conductors. Hence, if such layers are cooled to low temperatures, as is sometimes necessary in order to minimize the dark current, the resistivity of the cathode material may rise to such a high value that the photo-current can no longer flow. Even at normal temperatures difficulties may arise if very large currents are required to be taken. In such circumstances the difficulty can be overcome by depositing the antimony layer on a transparent highly-conducting backing layer.

As usual, any one of the alkali metals may be used instead of caesium, the result being a shift of the whole response curve towards the blue end of the spectrum as those of lower atomic number are used, and an overall reduction in efficiency. Moreover, bismuth, arsenic, thallium, gold or lead may be used instead of antimony, but the sensitivities are so poor as to make them of academic interest only; however, the Bi–Cs cathode is interesting in that it has a colour response which is relatively much stronger in the red end, and less in the blue end, of the spectrum than the Sb–Cs cathode.

(5.5.2) The Bi–Ag–Cs Cathode.

The difficulty of giving convincing physical reasons for the behaviour of photo-cathodes is well illustrated by this case. The basic layer is formed by evaporating (as before) a layer, either opaque or transparent, of an alloy of equal parts of bismuth and silver. For the transparent surface the thickness must be fairly carefully controlled, but it seems to be immaterial whether the two metal components are evaporated together or sequentially.

The next step is to oxidize the layer, and while this may be done by means of a discharge, as for Ag–O–Cs, it is preferable to do so by admitting about 0·1 mm Hg of oxygen and baking the tube at 200° C for a few minutes. The oxygen is then pumped away, caesium is admitted and the activation process is carried out almost exactly as for the Sb–Cs surface, except that the final oxidizing process is omitted.

The sensitivity of such a photocell is comparable with that of an Sb–Cs cell but rather lower. However, the important feature of this photo-cathode is that it has a lower response in the blue-violet range and a higher response in red-orange range. Its threshold wavelength is about 7 500 Å and consequently it has a rather higher dark current than the Sb–Cs surface at the same temperature. The colour response curve, shown as (*c*), Fig. 5, is the closest approach of any of the efficient photo-cathodes to that of the human eye. It can easily be corrected by means of a yellow (blue absorbing) light filter to fit the eye response curve, (*a*) in Fig. 5, fairly accurately.

It will be fairly obvious why a photocell with a colour response closely approximating to that of the human eye is so desirable. There are many cases in colorimetry where it is required to measure the intensity of light as seen by the human eye, but the most important case is in television, where the reproduced monochrome picture must be an acceptable reproduction of the original colour scene. Paradoxically, it is not so important in colour television, since independent controls are provided to give correct balance between the three colour components of the picture.

No attempt has yet been made to propound a theory to account for the sensitivity of this surface, but the following points should be noted:

The silver would not be oxidized as a result of the baking in oxygen, and hence the layer must be a mixture of BiO and metallic silver. The layer is a poor conductor, the conductivity being due primarily to the silver content. The caesium probably reacts with the BiO to form Cs_2O and free bismuth, or an intermetallic compound of bismuth and caesium. Finally a layer of free caesium is undoubtedly absorbed on, and in, the layer. It is a relatively poor conductor, and a better result is attained if it is formed on a conducting transparent surface on glass. Such a surface layer known by the trade name of Nesa, may be formed by spraying on to the glass, while it is held at a temperature near its softening point, a finely divided solution of stannous chloride in ethyl alcohol, water and hydrochloric acid.

(5.5.3) Composite Alloy Cathode.

A new type of photo-cathode has recently been described[15] which appears to be closely related to the intermetallic compound cathodes. Precise details have not been published, but it appears to be formed by evaporating a base layer of antimony and then admitting to the tube and baking successively, small quantities of sodium, potassium, and finally caesium.

Sensitivities reaching 200 to 300 μA/lumen to standard white light have been claimed, which represent a considerable increase in efficiency over the best photo-sensitive surfaces hitherto available. The increase is only about 50% in the blue end of the spectrum, compared with Sb–Cs cathodes, but the response appears to be quite efficient in the red end, and falls to zero only in the near infra-red. Such an increase in sensitivity to visible light would make possible a substantial increase in the sensitivity of many light-detecting and image-recording devices. However, no account has yet been published of its use in such devices.

(6) GENERAL CHARACTERISTICS OF PHOTOCELLS

Those characteristics which are common to all vacuum photocells will be discussed in the following Sections. Those which are more specifically characteristics of particular photo-cathodes are dealt with in the Sections dealing with those surfaces.

(6.1) Fatigue

This is the term applied to the fall in photo-sensitivity during operation of a photocell, from which it recovers on resting or removal of the incident light and applied electric field. It does not occur when either light or field is applied alone and it can be reduced by warming the photo-sensitive surface or irradiating it with infra-red light. These facts suggest that it may be due to depletion of the semi-conducting layer of electrons and/or the trapping of electrons by the oxygen atoms of the lattice, thus creating a space charge which screens the field from the base metal and so prevents further emission of electrons from it.

Effects which are more or less permanent and hence should not, strictly speaking, be called fatigue can be caused by the use of excessive light with the electric field applied and may be

due to some form of electrolysis through the layer. Furthermore, gas-filled cells are very subject to such effects, probably owing to the bombardment of the cathode by positive ions. Hence, high currents and high applied voltages tend to reduce the useful life of such cells.

The Sb–Cs and Ag–Bi–Cs cathodes are more susceptible to fatigue, both temporary and permanent, than the Ag–O–Cs cathodes, and approximate upper limits for current loading are $0{\cdot}1\,\mu A/cm^2$ for the former and $0{\cdot}5\,\mu A/cm^2$ for the latter.

(6.2) Angular and Velocity Distribution of Photo-Electrons

In any attempt to use photo-electrons, and especially if one wishes to form an electron image of the emission from a surface, it is important to know both the angular and velocity distributions of the electrons as they leave the surface, since these will produce aberrations in the electron image.

Measurements made of the angular distribution of photo-electrons from clean alkali metal surfaces indicate that it follows Lambert's law,[16] i.e. the emission per unit solid angle in any direction is proportional to the cosine of the angle that this direction makes with the normal to the surface. This is illustrated in Fig. 6, where OA and OB represent the emission per unit solid

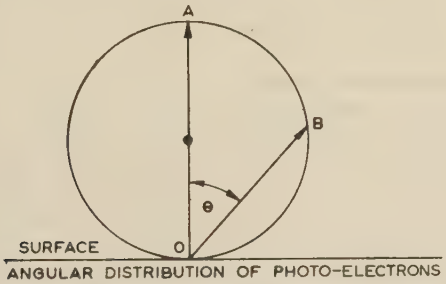


Fig. 6.—Polar diagram of angular distribution of photo-electrons.

angle normal to, and at an angle θ to the normal to, the surface respectively. This was found to hold for light of any wavelength or state of polarization at any angle of incidence, but so far as is known similar measurements have not been made on more modern photo-electric surfaces.

The velocity distribution of photo-electrons is proportional to the square root of the energy expressed in electron-volts, since $v = (2Ve/m)^{1/2}$. The maximum energy of the electrons depends on the wavelength of the exciting light and the work function of the surface, according to eqn. (5), and the distribution of energies lies between this maximum and zero, rising to a sharp maximum at about 80% of the maximum energy. Typical energy-distribution curves[17] for alakli metal surfaces irradiated with light of two different wavelengths, 4350 and 3650Å, are shown in Fig. 7.

It will be noticed that the difference between the maximum electron energies, and also between the energy values at which the maximum emission occurs, are very closely equal to the difference in the quantum energies of the incident photons. Electrons emitted from the more complex photo-cathodes show considerable irregularities in their energy distributions.

(6.3) Frequency Response

The time delay between the arrival of light on a photo-cathode and the liberation of an electron appears to be very small, probably less than 10^{-9} sec, and even smaller is the spread in this delay which would cause reduction in the modulation of the emitted electron current as compared with that of the incident light. Hence, for all practical purposes to date the emission of

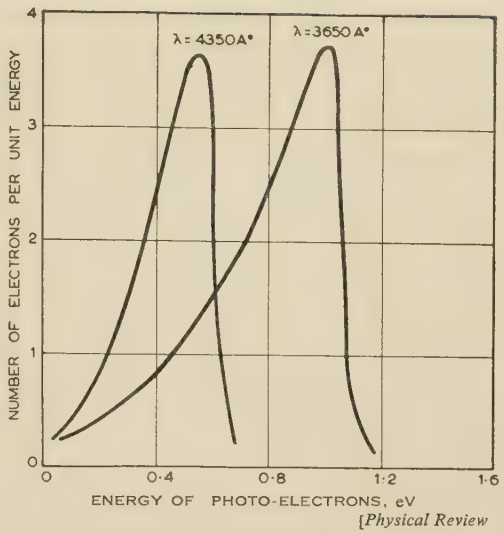


Fig. 7.—Energy distribution of photo-electrons liberated from a photo-cathode.

photo-electrons may be considered to be instantaneous with the arrival of light. However, the photo-electrons take an appreciable time to travel from the photo-cathode to the anode, and from the moment an electron leaves the cathode it begins to induce a charge on the anode. Thus, a pulse of electrons leaving the cathode in a time which is short compared with their time of flight to the anode will produce a pulse on the anode with a rise time comparable with the time of flight. Therefore, the limit to the frequency of light modulation that can be reproduced fully by a photocell is approximately the inverse of the time of flight of an electron from the cathode to the anode. If we consider the simplest case of plane parallel cathode and anode d cm apart and with a voltage V applied between them, the time of flight, T, will be $2d/v$, where v is the velocity of the electron on arrival at the anode and is given by

$$v = \left(\frac{2eV}{m}\right)^{1/2} = 6 \times 10^7 \sqrt{V} \text{ centimetres per second}$$

where V is in volts.

Hence, if $d = 2$cm and $V = 200$ volts, $T \simeq 5 \times 10^{-9}$ sec. Thus one would expect the frequency response to begin to fall off above 200 Mc/s. In practice, the effect is probably less than this in most tubes, where the induced effect of the electrons remains small during the greater part of their flight from cathode to anode.

More important are the effects of lack of geometrical symmetry in the tube, which results in different times of flight of the electrons from different parts of the cathode to the anode and also in the failure of some electrons to reach the anode on their first passage. They may follow a long orbit round the tube before being collected, and so contribute a long time-component to the induced charge on the anode.

It is possible to reduce the effect of the electrons in flight by placing between the cathode and anode a screening grid held a little below anode potential, so that the electrons travel only a short distance at high velocity after they are able to induce charges on the anode. This has the disadvantage of tending to increase the anode-earth capacitance, which reduces the efficiency of the input stage of an amplifier in which the tube may be included, especially at high frequencies. Nevertheless, this is substantially the arrangement used in some photo-multipliers, where the time of flight from cathode to anode may be much greater than the response time.

(6.4) Current/Voltage/Illumination Characteristics

The curves shown in Fig. 2 are obtained only with idealized photocells, and the type of curve obtained with practical cells, such as those shown in Fig. 4, are shown in Fig. 8. Each curve

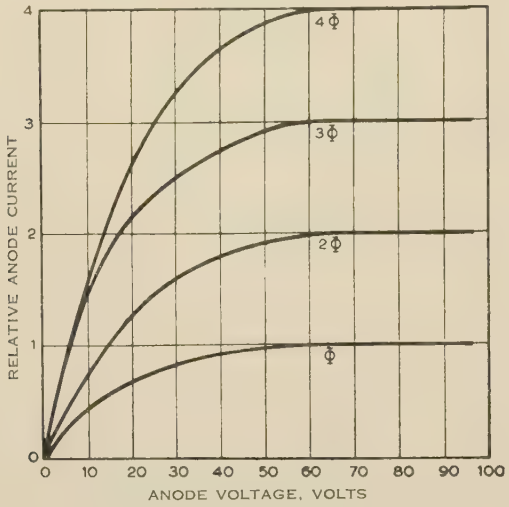


Fig. 8.—Current/voltage/illumination characteristics.

represents the way in which the anode current increases with increasing anode voltage, eventually reaching a substantially constant value between 50 and 100 volts. This saturation current represents the conditions under which the photo-electrons are substantially all collected on the anode.

If the illumination, Φ, on the cathode is doubled, trebled, etc., we get curves which are similar in shape, but in which the saturation current is very accurately proportional to the increase in light flux. This illustrates the very important feature of the photo-electric emission, namely that it is strictly proportional to the exciting light down to the smallest possible amounts.

The reason for the slow rise to saturation only after about 50 volts has been applied to the anode may be attributed to several causes, namely

(*a*) The surface, especially of an Ag–O–Cs cathode, may be rough or spongy; hence, some electrons emitted in 'valleys' collide with 'peaks' and are captured, thus failing to leave the surface. A strong field will penetrate into these valleys, pull out the electrons and accelerate them to the anode.

(*b*) Weak magnetic fields (e.g. that of the earth) may deflect slow electrons into circular orbits and return them to the parent surface. This effect is more pronounced on a rough surface. A strong electric field assists their escape.

(*c*) Incomplete collection of electrons on the anode. This may be due to the initial energies of some of the electrons or to bad design of the cell, so that some photo-electrons miss the anode until its voltage is increased sufficiently to attract all the electrons to it. Some may strike the glass walls of the tube and release secondary electrons, which then go to the anode.

(7) GAS-FILLED PHOTOCELLS

If a small quantity of gas is introduced into a photocell, some of the photo-electrons will collide with gas atoms on their way from the cathode to the anode; and if they have acquired an energy equal to, or greater than, the ionization potential of the gas atoms, ionization may occur and a free electron may be liberated from the atom to travel on towards the anode, while the residual positive ion will drift more slowly back to the cathode. The two free electrons may acquire sufficient energy for each to repeat the process, the resulting four electrons repeat it again, and so on until all reach the anode. The positive ions drift back to the cathode, where they also may liberate electrons. Hence, under the right conditions, a considerable amplification of the original photo-electric current can be achieved. This gain in sensitivity is of great importance in many practical applications of photocells, such as the reproduction of sound from film sound-tracks.

As has been noted, the alkali metals, especially caesium, used in activation of most photocells are extremely active elements chemically. Hence the only gases that can be admitted to such a photocell containing caesium without reactions that will be deleterious are the rare gases, i.e. helium, neon, argon, krypton and xenon.

Taken in this order, the atoms of these gases have increasing atomic weights and atomic radii and decreasing ionization potentials. It is desirable to have as low an atomic weight as possible, since the lighter the atom the faster will be the rate of diffusion of the positive ions back to the cathode. Conversely, the greater the radius of the atom the greater is the probability of an electron making an ionizing collision, which is also to be desired. Finally, the atoms should have as low an ionization potential as possible, since for a given potential difference this will result in more ions being formed. It is found that argon is the best compromise between these conflicting desiderata: it has an atomic weight of 40, an atomic radius of 1·8 Å and an ionization potential of 15·7 eV.

If the gas pressure in a photocell is very low, so that an original photo-electron makes on the average only one collision with a gas atom on its passage from the cathode to the anode, it follows that the amplification cannot exceed about 100%. Moreover, with the normal p.d. of about 100 volts between cathode and anode, some of the electrons will have acquired much greater energies than that required to ionize an atom, i.e. 15·7 eV, which will not be utilized. On the other hand, if the gas pressure is too high, electrons will make very many collisions before they have acquired sufficient energy from the field to be able to ionize an atom. Some of this energy is dissipated in inelastic, but not ionizing, collisions with gas atoms. Any ions produced drift back to the cathode so slowly that there is a high probability that they will recombine with electrons and so reduce the amplified current. Thus it is fairly obvious that there must be an optimum gas pressure at which, for a given electric field and tube geometry, the best amplification will be obtained.

It is a reasonable assumption that the optimum gas pressure will be such that in one mean free path an electron will acquire from the electric field an energy equal to the ionization potential of the gas. On this basis it can be calculated from gas-kinetic theory that the optimum pressure should be proportional to the field strength and inversely proportional to the ionization potential of the gas. The actual value works out to be of the order of a few tenths of a millimetre of mercury, which is in good agreement with the value found experimentally. For a given cell and applied voltage the variation in gas amplification with gas pressure is shown in Fig. 9. Clearly, in high vacuum the amplification is unity or there is zero addition to the primary photo-current. As the pressure is increased the amplification increases, passes through a maximum and then decreases. The maximum useful value is usually about a factor of ten; larger values can be attained, but the operation of the cell then tends to be somewhat unstable as the condition of complete instability is approached.

(7.1) Instability of Gas-Filled Photocells

The condition of instability is reached when, as the voltage is increased, the current becomes independent of the illumination

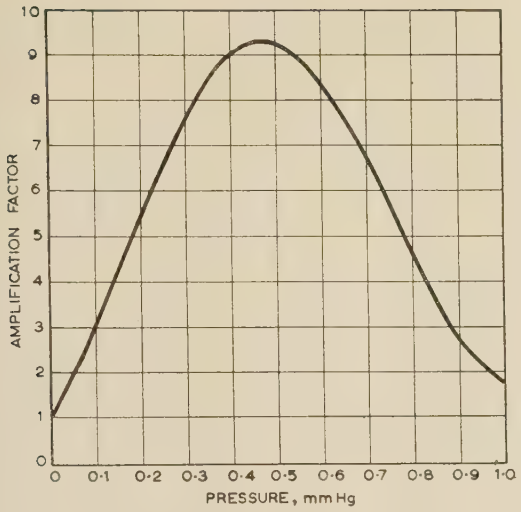


Fig. 9.—Variation of gas amplification with gas pressure.

and will continue when the light is cut off. This is shown in Fig. 10. Curve (*a*) is the saturation curve obtained from the cell under high vacuum, and curve (*b*) shows the increase in current relative to the saturated primary electron current as the voltage applied to the gas-filled tube is increased. This increase is approxi-

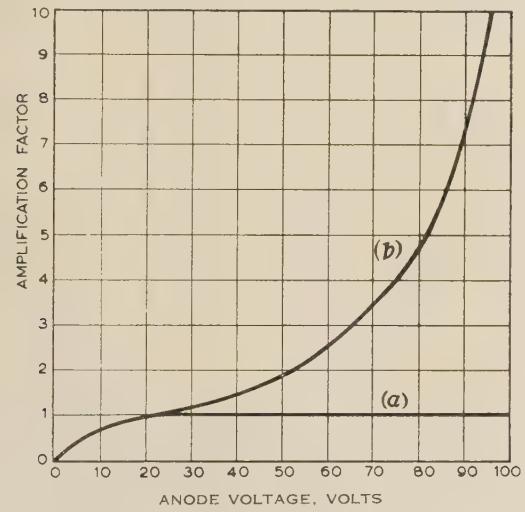


Fig. 10.—Current/voltage characteristics.
(*a*) Vacuum cell.
(*b*) Gas-filled cell.

mately linear over a large range of voltage and then, as the voltage is increased further, it begins to rise until its slope becomes almost vertical. This rapid rise cannot be explained by increased ionization in the gas, and an additional mechanism must be introduced to account for it; this is found to be the liberation of electrons from the surface of the cathode by the positive ions formed by the electrons. If each photo-electron results in n electrons reaching the anode, there will have been $n-1$ ionizing collisions and $n-1$ positive ions will return to the cathode. If, on the average, these $n-1$ ions produce at least one further electron, the discharge will be self-maintaining and the condition of instability will result. Thus, if each positive ion produces τ electrons at the cathode, the condition for instability is given by

$$(n-1)\tau \geqslant 1$$

(7.2) Current/Light Characteristics

In gas-filled cells the relationship between current and light flux is not linear, as it is in a vacuum cell. This is attributed to a lowering of the work function of the cathode surface by the positive ions which return to it and are believed to remain unneutralized upon it for a considerable time, thus assisting the liberation of electrons both by photons and ions. This non-linearity is not sufficient to cause serious error, except in measurements of great accuracy.

(7.3) Dynamic Characteristics

It has been shown that in a vacuum photocell the rapidity of response in the current output to the light input is determined by the time taken by the photo-electrons to travel from the cathode to the anode. In a gas-filled cell the positive ions will also contribute to the response, but they move much more slowly than electrons, by a ratio proportional to the square root of their relative masses, so that their contribution to the response will be reduced in rapidity—for argon, compared with electrons, by the factor $(40 \times 1\,823)^{1/2} = 271$. Two other effects also contribute: the positive ions may remain on the surface of the cathode for an appreciable time before giving up their charge, and metastable atoms are formed in the gas which give up their energy only when they reach a surface, their energy being used to liberate an electron from that surface. Being uncharged, they reach the cathode surface only by diffusion, and this may take a considerable time; hence they contribute a very long time-lag component to the response of the cell.

The combined effect of these three causes is that, on the almost instantaneous application of illumination, the photo-current will rise very rapidly to about 75% of its final value and then approach this value exponentially over a period of about 1 millisec. If the modulation of the output current is plotted as a function of the frequency of modulation of a constant input illumination, a curve of the type shown in Fig. 11 is obtained.

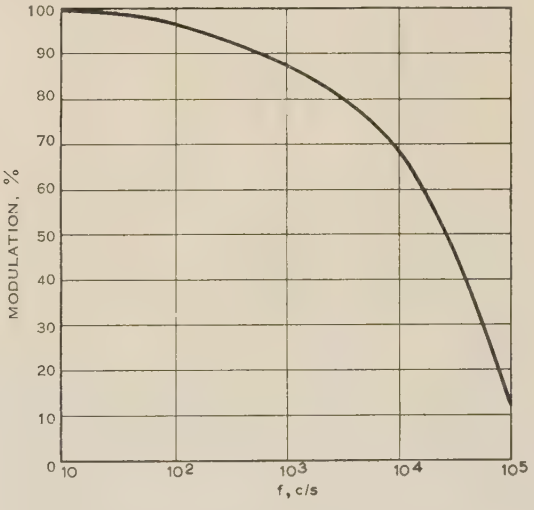


Fig. 11.—Variation of output modulation amplitude with light input frequency.

This shows a noticeable fall-off in amplitude at 5 kc/s, and this may be 25% at 10 kc/s and 50% at 20 kc/s. This is not important for many measurements, but it becomes significant at the top audio frequencies, where compensating peaked circuits must be used. At top video frequencies in television it makes the cell quite useless.

The intermetallic-compound cathodes cannot be used success-

fully in gas-filled photocells, since the bombardment of the cathode by the positive argon ions quickly destroys the efficiency of the surface. The silver-oxygen-caesium surface is much more resistant to this effect, but it is slowly destroyed and the effective life of the tube is determined by the ion bombardment.

(8) ELECTRON-MULTIPLIER PHOTO-CELLS

A method of achieving very large amplification of a primary photo-electric current which has not the disadvantages of gas amplification is provided by the phenomenon of secondary-electron emission.

If the primary photo-electrons are accelerated on to a suitable surface, each may liberate a number, δ ($\simeq 5$), of secondary electrons. If, in turn, these secondary electrons are accelerated on to another surface from which each liberates δ tertiary electrons, the total number will have become δ^2. If the process is repeated through n stages, each original electron will have been multiplied by a factor $M = \delta^n$. If $\delta = 5$ and $n = 10$, M will be about 10^7.

(8.1) Secondary-Electron Emission

The number of secondary electrons, δ, liberated by each primary depends on the material of the solid, the state of its surface, the energy of the primary electrons and the angle of incidence at which they reach the surface.

Three representative curves for (*a*) bad, (*b*) medium and (*c*) good secondary emitters are given in Fig. 12, showing the relationship between δ and the energy of the incident primary electrons. The bad emitters are materials with rough surfaces, such as carbon or aluminium oxide, on which secondary electrons tend to be trapped mechanically. Medium emitters are the very pure metal surfaces for which the maximum δ rises only a little above unity. All efficient photo-electric surfaces, intrinsic semiconductors and many insulators are good secondary emitters. Thus the Ag–O–Cs and Cs_3Sb photo-surfaces, the alkali halides and magnesium and beryllium oxides are good emitting surfaces, although the latter can be used effectively only if they can be made conducting by the admixture of finely divided metal, or in extremely thin layers.

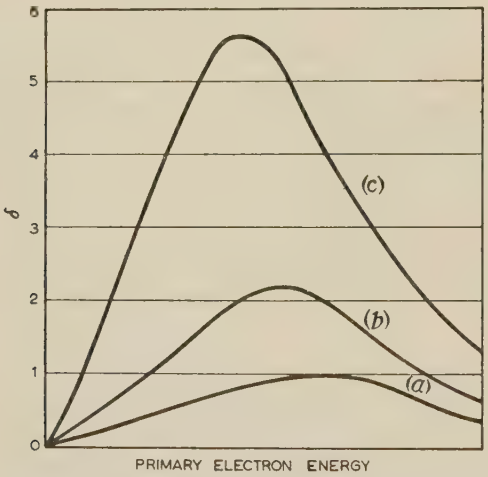


Fig. 12.—Secondary-electron emission as a function of primary-electron energy.

The characteristic shape of the curves in Fig. 12 is readily explained by observing that at low energy of the incident electrons they will not be able to impart enough energy to many electrons of the solid to enable them to escape from the surface. As its energy increases to 100 eV or so, the primary electron will produce a considerable number of excited electrons quite near the surface, and a fair proportion of these will escape. As the velocity of the incident electron increases, still more excited electrons are produced, but because a fast electron produces fewer ions per unit length of its path than a slow electron, we find that, although many more excited electrons are produced by the fast primary electron, a larger proportion are produced at such a depth beneath the surface that they fail to escape. Thus the curve passes through a maximum and gradually falls, eventually falling below $\delta = 1$ at high primary energies. The maximum value of δ is usually found at primary-electron energies of about 500 eV, when it may reach the value of 10; however, at energies of about 200 eV values of 6–8 are frequently obtained.

An important fact about secondary-electron emission is that there is an extremely short time-delay between the arrival of the primary and the liberation of the resulting secondary electrons at a surface. Thus, by directing the primary photo-electrons on to an efficient secondary-emitting surface, henceforth termed a 'dynode', and collecting the secondary electrons released on to an anode, an amplification of the primary photo-current by a factor of 5–10 can be achieved without any appreciable loss of frequency response.

(8.2) Single-Stage Photomultiplier

A simple single-stage photomultiplier is shown in Fig. 13. When a beam of light passes through the glass wall of the envelope and falls on the photo-cathode, photo-electrons, *e*, liberated from the cathode are accelerated towards the metal grid; this is held at about 250 volts positive relative to the cathode, but it has a low shadow ratio and hence most of the electrons are carried through it by their momentum to the dynode, whose potential is about 50 volts lower than that of the grid. Although the electrons lose energy as they pass from the grid to the dynode, they arrive at the latter with an energy of about 200 eV, which is sufficient to liberate 5–8 secondary electrons, *e'*, from its surface. These find themselves in an electric field which accelerates them to the grid, where they are collected. The multiplied current so collected is the 'signal current' which may produce a voltage signal in the load resistor R or be measured by an ammeter. Thus, a gain in current output per unit light input by a factor of between 5 and 10, with an applied overall voltage of 250 volts, can be achieved without impairment of response.

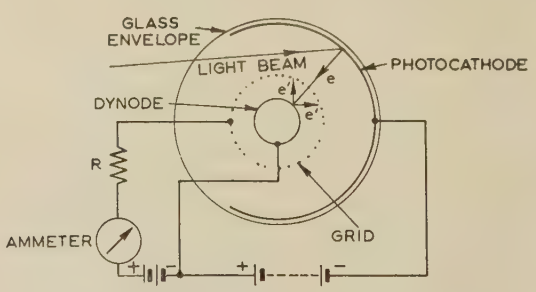


Fig. 13.—Simple single-stage photo-multiplier.

The essential feature of electron-multiplier design is illustrated in this tube, i.e. the primary electrons must be given sufficient energy to enable them to approach the surface of the multiplying dynode through a decelerating electric field and still have sufficient energy left to release a useful number of secondary electrons from the surface when they hit it. The secondary electrons then find themselves in a field that removes them from their parent

surface and carries them on to a collecting electrode or, in a multi-stage electron multiplier, on to the next dynode.

(8.3) Multi-Stage Photomultipliers

It was obvious at an early date that, if the multiplication process could be repeated n times and a useful multiplication ratio, k, achieved each time, the overall gain, M, for say $k = 5$ and $n = 10$ will be

$$M = n^k = 5^{10} \simeq 10^7$$

Such an enormous gain is achieved without appreciable loss of frequency response and hence enables minute photo-currents to be amplified to a level where they far exceed the noise current of any amplifier that might be required. Moreover, it was found that such a multiplier, if designed so as to avoid spurious dark-currents, gave almost the same signal/noise ratio at its output, S_A, as was inherent in the primary current at the photo-cathode, S_C. Experimental measurements support the deduction from statistical theory that

$$S_A = S_C \sqrt{\frac{k-1}{k}}$$

i.e. if $k = 5$, S_A is only 10% lower than S_C. This is a very low price to pay for such a large gain, and hence multi-stage photo-multipliers are now widely used in the measurement of low-intensity high-frequency light signals.

Space will not permit more than the briefest outline of multi-stage photomultipliers. However, three frequently used types of dynode structure are illustrated in Fig. 14. The early type in

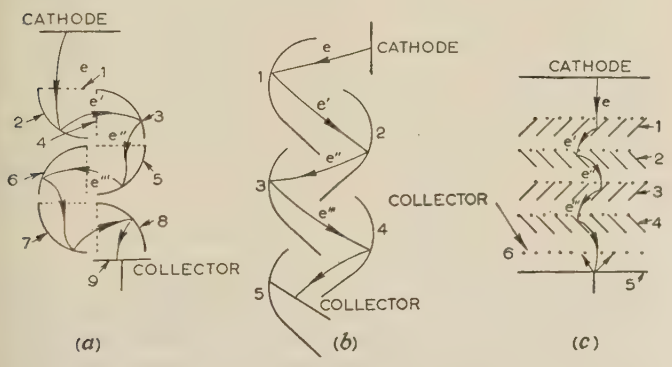


Fig. 14.—Three types of photo-multiplier dynode structure.
(*a*) Cage type.
(*b*) Focusing type.
(*c*) Venetian-blind type.

which a magnetic field was used to guide the electrons from one dynode to the next was soon discontinued, because it was inconvenient, and the type in which a series of metal meshes acted as dynodes proved to be very inefficient. Fig. 14(*a*) illustrates the cage-type multiplier;[18] each dynode is a quarter-cylindrical surface, with a low-shadow-ratio mesh, forming a second wall through which the electrons, *e*, enter with, say, 200 eV energy. The second dynode is similar, with a cylindrical surface and a mesh wall which is arranged to stretch across the open wall of the first dynode. It is held at about 200 volts relative to the first dynode, and hence secondary electrons, *e'*, liberated by a primary electron will be attracted towards the mesh and will, nearly all, pass through and impinge on the second electrode. The trajectories of the fast primary electrons are only very slightly deflected by the field between the first and the second dynode. A similar process occurs between the second and third, third and fourth, and so on, until the very greatly multiplied stream of electrons from the final dynode is collected on the anode or collector.

Another well-established multiplier structure[19] is shown in Fig. 14(*b*). In this case, photo-electrons, *e*, fall upon the first dynode and release secondary electrons; these are attracted to the second dynode, which is held at a higher positive potential than the first, from which they liberate secondary electrons. The dynodes are so shaped that the field gradient at the surface of the second accelerates the secondary electrons towards the third; similarly, secondary electrons liberated from the third are accelerated to the fourth, and so on.

In this design the electrodes are so shaped that the screening meshes used in the first example can be omitted. Also, the paths of successive batches of secondary electrons cross over and so average out the total path length of electrons through the multiplier. Hence, electrons liberated by a short flash of light on the photo-cathode will produce a very short output pulse of electrons at the collector electrode.

A third, well known, design of multiplier dynode[20] is shown in Fig. 14(*c*), and is known as the 'Venetian blind' type. Each dynode is in the form of a Venetian blind with the slats at an angle of about 45° to the plane of the blind, and with a fine metal mesh fixed to the surface of the blind upon which the primary electrons, *e*, impinge. Because of the slight overlap of the slats in the first dynode, few primary electrons can pass through without impinging on a surface from which they will liberate secondary electrons. These will be accelerated towards the next dynode, and are protected from the suppressing field from the previous electrode by the screening mesh. Most of the secondary electrons released at the surface of one dynode can be attracted to the next dynode, but a few that are released in the corner between the slat and the screening mesh are in a very weak field and may either not escape at all or only do so slowly. Thus a group of electrons entering the first dynode together will produce an output pulse which is somewhat lengthened in duration. After an adequate number of multiplying processes, say 10–15, the output current may be between 10^6 and 10^{10} times the input current. A practical limit tends to be imposed by space charge in the last few stages, which makes difficult the collection of a large secondary current from one dynode to the next, and the output current ceases to be linear relative to the input light.

(9) IMAGE CONVERTORS

Image convertors are photo-electric tubes used to convert an optical image of light of one colour, e.g. infra-red or ultra-violet, to light of another colour, generally in the visible range. They may also be used for image intensification, and they are integral components of many television pick-up tubes. Their operation depends on two phenomena: first, that the photo-electrons released from a photo-cathode can be accelerated to high kinetic energy by electric fields, and that this energy can be converted to light of a chosen colour by directing the electrons on to a fluorescent screen; second, the photo-electrons liberated from a photo-cathode may be focused by electrostatic or electro-magnetic lenses to form an electron image on a fluorescent screen and hence the light emission they excite forms a visible image.

A simple image convertor is illustrated in Fig. 15. It consists of a highly evacuated glass envelope with a transparent photo-cathode on a window at one end and a fluorescent screen on the window at the opposite end. A series of ring electrodes on the wall of the tube are held at progressively higher potentials, increasing from the cathode to the anode, by a voltage divider. Light focused on to the photo-cathode by the lens liberates photo-electrons, which are accelerated by this field gradient towards

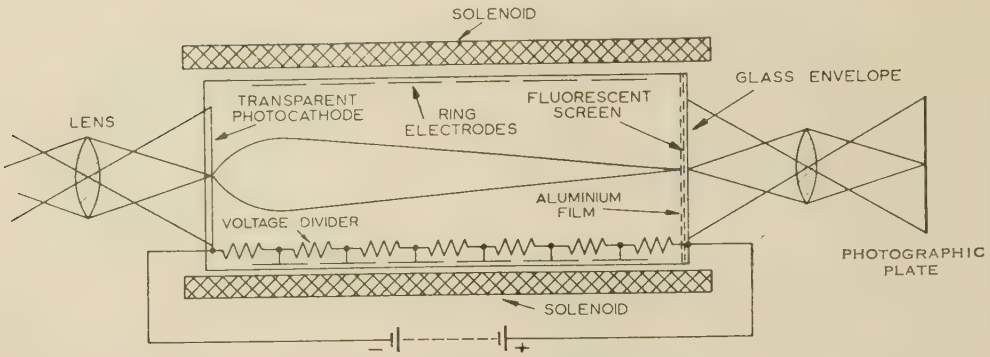


Fig. 15.—Photo-electric image convertor.

the screen, and they are focused by the axial magnetic field produced by the solenoid to form an electron image on the fluorescent screen. The electron image in this case is the same size as the optical image and is not rotated. If short magnetic coils and non-uniform accelerating fields are used, magnification or demagnification as well as rotation of the electron image can be obtained. Again, electrostatic electron lenses or a combination of the two may be used.

If average white light falls on an efficient photo-cathode, one photon in about 25 will liberate one photo-electron. When such a photo-electron has been accelerated to 20 keV it is capable of liberating about 1 000 photons from a suitable fluorescent screen. Hence an overall gain of 40 is theoretically possible. However, this apparently large intensification factor is to some extent illusory, since the light from the fluorescent image is isotropic and hence in visual observation, or in focusing an image of it on a photographic plate, such a small percentage of it is collected and used that the overall advantage is usually small, if not negative.

(10) TELEVISION PICK-UP TUBES

The photo-electric effect is the fundamental basis of the generation of television signals, since this depends on the conversion of light energy into electrical energy. A detailed account of this branch of photo-electronics is beyond the scope of this review and has been given elsewhere.[21] However, since devices of this type are now probably the most important of all photo-electric tubes they cannot be completely omitted from this survey.

The basic principle of all such television signal-generating devices is that the light distribution of the optical image is sampled point by point in an orderly, predetermined manner, termed 'scanning', and each sample, averaged over a predetermined time interval, is measured by passing the light into a photo-electric cell and measuring the average current passed. The variable current thus produced in scanning the picture elements constitutes the picture signal.

The optical image may be scanned by an aperture corresponding in area to one picture point, the pencil of light passing through the aperture being directed into a photocell or, more appropriately, a photo-multiplier such as those described in Section 8.3. Alternatively, the optical image may be formed on a photo-electric cathode and the emitted photo-electrons may be accelerated and focused to form an electron image as described in Section 9. The electron image may then be scanned over a limiting aperture corresponding to one picture point, so that the electrons which pass through it pass into an electron-multiplier system. The output of the electron multiplier is the picture signal current.

In both these cases the light of the optical image is sampled for the time required to scan one picture point ($\sim 2 \times 10^{-7}$ sec). For the rest of the time the light from each picture point is not used, and this results in a serious loss of information. It will be evident that, if the sampling period can be extended to approximately the picture frame period ($\frac{1}{25}$ sec), an enormous increase in sensitivity can be achieved—equal to the number of picture points in a picture, e.g. $\sim 10^5$—at the expense of a certain amount of blurring of image definition resulting from any movement taking place in $\frac{1}{25}$ sec. In order to do this it is necessary to integrate the photo-electric charges liberated by light, or charges produced by them, during one frame period on a 2-dimensional array of minute condensers, all of which have one plate in common. These condensers are then scanned by a beam of electrons, which discharges them in sequence and thus induces a fluctuating picture signal from the common plate (signal plate) to earth through the signal resistance.

A television signal-generating tube is shown diagrammatically in Fig. 16. It consists of a highly evacuated glass envelope with a simple electron gun at one end from which a beam of electrons is projected on to the target at the other end. This electron beam is focused by the field of a solenoid coil and scanned in a raster over the surface of the target by two pairs of saddle coils. The target consists of a thin sheet of glass or mica with a transparent conducting coating on the surface turned away from the scanning beam, and connected to an amplifier and to earth through the signal resistor, R. The other surface of the dielectric sheet is covered by a vast number of minute 'islands' of transparent photo-sensitive material, each of which is highly insulated from its neighbours. The optical image is formed by the lens through the end window of the glass envelope, through the conducting signal plate, transparent dielectric and transparent photo-electric element. Photo-electrons are liberated from the free surface of each of these elements, and are removed by an electric field to the walls of the tube. This loss of electrons leaves each cell charged positively, and this goes on continuously, the charge on each mosaic element building up throughout a complete frame period and then being suddenly discharged to a constant datum potential at each sweep of the scanning beam.

Alternatively, the optical image may be formed on a conducting transparent photo-surface and the emitted photo-electrons then accelerated and focused as illustrated in the image convertor (Fig. 15) on to a target. The target in this case is similar to that shown in Fig. 16; however, the mosaic elements are not photo-sensitive, but are good secondary-electron emitters. The positive charges are built up by secondary-electron emission from the mosaic excited by the primary photo-electrons. The charged mosaic elements are then discharged by an electron beam, as before, to generate the picture signal.

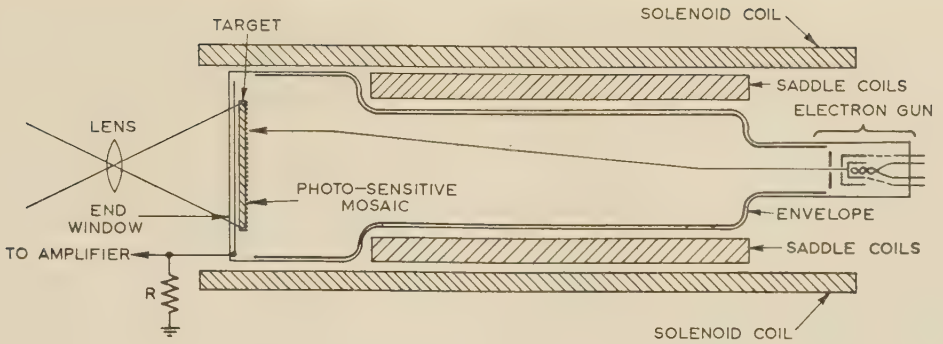


Fig. 16.—Diagrammatic television pick-up tube.

Again, the target, as shown in Fig. 16, may comprise a conducting signal-plate upon which is deposited a thin layer of photo-conductive highly insulating material, such as amorphous selenium. This layer takes the place of both the dielectric and the photo-emissive mosaic of the first example. A potential difference is established across this layer, and consequently the greater the light intensity at any point the more current flows through it at that point. Hence a distribution of charges corresponding to the distribution of light in the optical image builds up on the free surface of the photo-conductive layer. This is discharged as before by the scanning beam to produce the picture-signal current.

(11) INTERNAL PHOTO-ELECTRIC EFFECT

Electrons which gain energy by absorbing photons but do not escape from the solid body in which they originate may produce electrical effects—either an increase in conductivity, which is termed photo-conductivity, or a voltage difference, which is known as the photo-voltaic effect. A full treatment of this subject should deal with the solid-state theory of insulators, conductors and semi-conductors,[22] but it is possible to sketch this theory only very briefly in this review.

(11.1) Photo-Conductivity

All photo-conductors are also semi-conductors, and may be divided into three classes, namely

(*a*) *Intrinsic semi-conductors.*—These are pure substances in which the top energy level of the normally-filled band is separated by a very narrow energy gap from the normally-empty conduction band. In this case the electrons from the filled band may be excited by thermal energy or by absorption of a photon, and raised to the conduction band. The relative mobilities of the excited electron and the resulting 'hole', both of which contribute to conductivity, will determine whether the substance acts as an *n*-type or a *p*-type photo-conductor.

(*b*) *Impurity semi-conductors, n-type (donor) or excess type.*—In this case atoms of impurities in the crystal lattice produce distortions in the energy levels such that the top occupied electron level lies only a small energy step below the empty conduction band of the lattice. Electrons may be raised across this step occasionally by thermal excitation. A photon can cause such a transition if $hf > Ve$, where V is the energy step expressed in volts. Once an electron is lifted to the conduction band it will drift under the influence of an applied electric field until it is trapped by another crystal fault. Conduction in this case is by electrons only.

(*c*) *Impurity semi-conductors, p-type (acceptor) or defect type.*—In this case the lowest unfilled energy level, resulting from the distortions introduced by the foreign atoms—impurities—in the lattice lies slightly above the top filled band of the lattice. Thus, one of the electrons of the filled band may be excited by a photon to the empty impurity level, leaving an electron deficiency, or hole, in the normally filled band which can drift under the influence of an applied field towards the cathode. Eventually it may be trapped by an electron from another impurity state falling into it or it may be neutralized by an electron from the cathode. In this case conduction is due to holes only.

If the excitation of electrons and holes were the only cause of photo-conductivity, the current would soon stop, owing to negative and positive space charges near the anode and cathode respectively where electrons and holes had been trapped. Moreover, the photo-current would cease, within a time comparable with the very short 'life' of a hole or an electron, after the removal of the light. In fact, the current does not decrease, but is enhanced, and it may continue for a considerable time after the removal of the light. This is due to secondary effects, namely the release of trapped electrons and holes by thermal excitation and the extraction of additional holes from the anode and electrons from the cathode by the space charge set up within the photo-conductor. These two effects account for the fact that more than one electron is released per photon, owing to the internal multiplying action, and that there is a considerable time lag in the phenomenon. A very large number of substances exhibit photo-conductivity to some extent. Of these the most important are: sulphides, selenides, oxides, halides and the transition elements such as selenium.

(11.2) The Selenium Cell

This is the best known of the photo-conductive cells and it has been in use for a very long time. However, it is now obsolescent and is being replaced by the more efficient 'thalophide' cell.

The cell is formed by applying a thin layer of selenium to bridge the gap between two electrodes in the form of interlocking combs on an insulating surface, as shown in Fig. 17; this gives a long border between the two electrodes. The selenium layer, shown as a hatched area in the diagram, may be spread on the surface in the molten form or evaporated on to it *in vacuo* in the highly insulating amorphous form. It must then be annealed

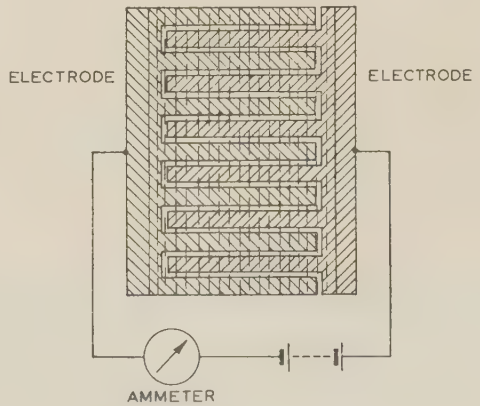


Fig. 17.—Interdigital structure of photo-conductive cell.

at about 100° C to convert it to the grey monoclinic crystalline form which exhibits photo-conductivity. In operation a potential difference is applied between the electrodes from a battery, and the circulating current is measured by the meter.

The characteristics of such a cell depend very much on the exact mode of preparation, but in all cases it shows an appreciable time lag, which indicates the very important part played in the conduction by secondary effects. There is, in general, a rapid, partial, response due to light fluctuations, attributed to the primary photo-conduction, followed by a slow, residual, response caused by the secondary effects. Thus, although there is a maximum response to slow light changes, there is still some useful response at several kilocycles per second.

The dark resistance of such cells ranges from about 10^5 to 10^7 ohms, the operating voltage being roughly in the same proportion, from 10 to 500 volts respectively. The higher the voltage applied to a cell, the larger the current that flows and so the larger the sensitivity. However, this cannot be pushed too far, since the cell tends to break down under intense fields applied across the border and undue noise results. The sensitivity in this case is measured by the change in resistance with incident light. A typical cell may fall to 25% of its dark resistance for 0·1 lumen incident light. The colour response is fairly uniform throughout the visible spectrum and extends to about 8 000 Å. Furthermore, the photo-current is not strictly proportional to the incident light, but falls below proportionality as the light is increased.

(11.3) The Amorphous Selenium Cell

It was only comparatively recently that the photo-conductive nature of amorphous selenium was recognized and investigated. This was probably due to the fact that its very high resistivity in the dark makes it suitable for use in television pick-up tubes. Amorphous selenium is obtained by evaporating the material on to a cold surface *in vacuo*, when it appears as a ruby-red film by transmitted light and has a resistivity quoted by different authors as between 10^{12} and 10^{15} ohm-cm. It is, in fact, a good insulator, which makes it unsuitable for use in the interlocking comb-electrode type of cell, but it is an essential characteristic for photo-conductors employed in television pick-up tubes. One might imagine that this form of selenium might be used to make a photocell by forming a film of it on a transparent conducting glass sheet and then applying, by evaporation or sputtering, a transparent metal film to the other side. Through these contacts a voltage could be applied across the layer and light could fall on it through the transparent electrodes. This is what is done, in principle, in photo-conductive television pick-up tubes, contact being made to the thin selenium layer by the transparent conducting signal plate on one surface and by the scanning electron beam on the other. A sensitivity of several hundred microamperes per lumen can be achieved. However, for practical light-meters it has two disadvantages: (*a*) the film almost always has minute pin-holes, through which short-circuits occur and give rise to excessive dark currents; (*b*) the amorphous selenium is slightly unstable even at room temperature and tends to revert slowly to the crystalline semi-conducting form in small areas. This again results in increased dark current.

(11.4) The Thalophide Cell[23]

This is now the most efficient photo-conductive cell in the visible and near infra-red wavelength range. It is prepared by evaporating a thin film of thallous sulphide (Tl_2S) about 0·5 μ thick on to an interlocking-comb electrode system (see Fig. 17) as used for the selenium cell. The layer is then baked at about 300° C in a small pressure (~1 mm Hg) of oxygen for a few minutes; this converts the thallous sulphide to thallous oxysulphide (Tl_2SO_2), yellow in colour and of much greater resistivity and higher photo-sensitivity than the original layer.

As with the selenium cell, the dark current and photo-current are both proportional to the applied voltage, but the increase of conductivity (or the photo-current) is not proportional to the incident light, the cell becoming less efficient with increasing light. The frequency response is poor, falling rapidly above about 100 c/s, although this improves with increasing temperature while the sensitivity decreases. The colour response is fairly uniform throughout the visible region, but rises slightly from blue to red, reaching a maximum at about 9 000 Å and falling to the threshold of response at 13 000 Å.

The mechanism of photo-conduction and the effect of the oxidation is worth noting.[24] The thallous sulphide is a rather inefficient *n*-type semi-conductor, the mobility of the electrons exceeding that of the holes. However, after oxidation the electrons tend to be trapped by adsorbed oxygen atoms and the layer becomes a *p*-type semi-conductor. Moreover, the negative oxygen ions have very low mobility, and hence will neutralize the space charge due to a much larger current for a given applied voltage. This is exactly analogous to the effect of positive ions in neutralizing space charge due to a much larger fast-moving electron current, as for example in a gas-filled diode.

(11.5) The Lead Sulphide Cell (and others)[25]

This again is prepared in the same general way as the thalophide cell and is activated by oxidation. Because of its low resistivity it is only necessary to employ quite a short gap bridged by the lead sulphide. The special importance of this material arises from its response, which stretches far into the infra-red, the threshold wavelength being at about 3·5 μ and its peak response between 2 and 3 μ, depending on the temperature. Thus it is very efficient for detecting infra-red radiation of long wavelengths which can penetrate cloud and fog, and was therefore of great importance for invisible signalling and other war-time operations. For these purposes the cells are operated at low temperatures, that of carbon-dioxide snow (−80° C) or even liquid air (−187° C), in order to minimize the dark current.

Lead selenide and lead telluride have useful sensitivity to even longer wavelengths, the thresholds increasing with decreasing temperature. Thus, lead telluride has threshold wavelengths of about 4, 5 and 6 μ at room, liquid-air and liquid-hydrogen temperatures, respectively, while lead selenide has corresponding threholds of about 5, 7 and 8 μ. For cells operating at such long wavelengths, a window of sapphire (Al_2O_6) or periclase (MgO) must be provided, and the photo-conductive layer is formed on a surface that can be cooled by a refrigerant.

(11.6) Other Photoconductors

Many other semi-conductors are also photo-conductors, e.g. cadmium sulphide, germanium and silicon and intermetallic compounds. The photo-conductive characteristics of silicon and germanium have been investigated very thoroughly, and it has been found that they may exhibit photo-conductive effects either as pure metals, i.e. as intrinsic semi-conductors, or as 'doped' metals which are either *n*-type or *p*-type semi-conductors.

Space does not allow an adequate discussion of the complex physics of these devices beyond the brief outline already given in Section 11.1, and the reader is referred to a comprehensive treatment[22,31] of the subject for more details.

Briefly summarizing it may be said that these photo-conductors usually exhibit high quantum efficiency, high dark current, lack of response to high frequencies and a response that extends far into the infra-red wavelengths. They may be expected to be of great importance in the future in many aspects of light and image detection.

(11.7) The Photo-Voltaic Effect

If a copper plate is oxidized by baking in air to about 1 000° C, a layer of red cuprous oxide (Cu_2O) will form on its surface covered by a layer of black cupric oxide (CuO). If the cuprous oxide is polished or dissolved off (using a solution of sodium cyanide) and a plate of lead is pressed against the cupric oxide surface, a very efficient rectifier is formed; electron current passes readily under an applied voltage from the copper to the lead, but against a high resistance in the reverse direction. It was soon found[26–28] that, if a transparent conducting electrode is applied to the cuprous-oxide surface, illumination falling on the semi-conducting layer can cause a current to flow in an external circuit without the application of an e.m.f. from a battery—hence the term 'photo-voltaic effect', since the light produces a potential difference across the interface between the copper and the cuprous oxide.

Clearly, in this type of cell the light must traverse the cuprous-oxide layer before reaching the interface where the photo-electric action occurs, and for this reason it is termed a 'back wall' cell. This is naturally inefficient, since much of the light—especially in the visible region—is absorbed in the cuprous oxide. Lange found that if a transparent contact electrode were applied to the cuprous oxide layer by sputtering on gold, platinum, nickel, etc., in a reducing atmosphere the photo-voltaic effect appeared between this layer and the cuprous oxide. Under these conditions the light is much more efficiently used and the cell is termed a 'front wall' cell.[29] It is to be noted that the interface between the transparent contact electrode and the cuprous oxide is effective only when formed in this way, and the effect is not obtained if the gold is evaporated in vacuum. It appears that it is necessary to sputter the gold to get very intimate contact and to use a reducing atmosphere to form a thin layer of oxygen-free cuprous oxide and possibly metallic copper. In each case the effect of light is to cause electrons to pass from the barrier layer (*Sperrschicht*) into the adjacent conductor, i.e. in the high-resistance direction.

(11.8) The Origin of the Photo-Voltaic Effect[30]

Cuprous oxide containing an excess of oxygen is a *p*-type semi-conductor. Excited electrons raised from the filled bands to fill impurity levels provided by the absorbed oxygen atoms require at least 0·3 eV energy, but they require 0·7 eV to reach the empty conduction band.

If we consider the interface between the copper and the cuprous oxide, since the latter has a much higher work function than the former, electrons will pass from the copper to fill the impurity levels provided by the oxygen atoms in the cuprous oxide. This builds up a space charge, which is negative on the oxide side and positive on the pure-metal side, resulting in the potential diagram shown in Fig. 18.

The field and the thickness of the barrier layer are such that the top of the Fermi limit of electron-energy distribution in the metal lies midway between the top of the filled bands and the impurity (acceptor) levels in the interior of the semi-conductor. At the metal–semi-conductor interface the levels are determined by the two work functions. Equilibrium exists under these conditions and there is no drift of charges in either direction.

If now electrons are excited by photons and lifted from the filled band to the conduction band in the body of the semi-conductor they will rapidly fall back into an impurity level and so do not produce any potential difference. If, however, excitation takes place in the barrier layer where the impurity levels have been filled by electrons from the metal, there is no empty impurity level for the excited electron to fall into, and so it has a long free life in the conduction band during which the field

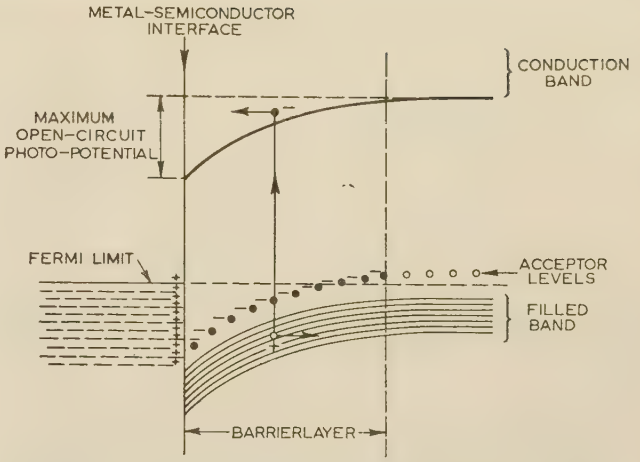


Fig. 18.—Energy levels at barrier layer of photo-voltaic cell.

across the interface will accelerate it towards the copper. The open-circuit p.d. should be equal to this accelerating field, which approaches the difference in the work functions of the two materials. This agrees well with observations.

(11.9) The Selenium Photo-Voltaic Cell

The selenium photo-voltaic cell (Fig. 19) is the only one of practical importance, and it is formed by spreading a thin layer

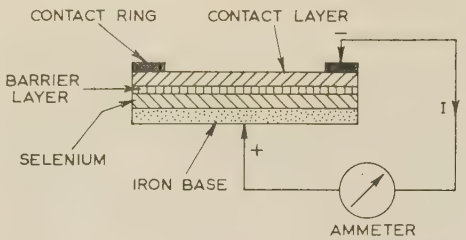


Fig. 19.—Cross-section of photo-voltaic cell.

of molten selenium on an iron base, at about 200° C and annealing at 80° C to produce the crystalline form. Small quantities of zirconium, thallium or cerium are sometimes added to the selenium to give greater efficiency. A transparent layer of metal is then evaporated, sputtered or sprayed on to the selenium surface to form a contact, the material and its method of application being unimportant. A robust metal ring is then applied to make contact with the transparent layer, and serves as the negative pole connected to the external circuit through the micro-ammeter. Such a cell is a front-wall cell and its mechanism of operation is probably similar to that of the copper-oxide cell, the selenium being also a *p*-type semi-conductor.

The colour response is close to that of the human eye, and hence it can easily be corrected by a suitable colour filter to give a very close match. This feature, together with the fact that no battery is required for its operation, results in the great value of such a cell for practical measurement of light, and especially as an exposure meter in photography.

The equivalent circuit of such a cell is as shown in Fig. 20. The cell, which itself includes all those circuit elements included within the dotted boundaries in the diagram, may be regarded as a current generator in series with a small internal resistance R_s and an external load R_L, but also shunted by the internal capacitance C and resistance R_i across the barrier layer. The current, I,

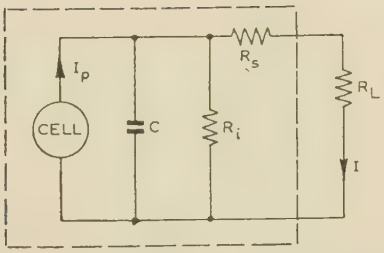


Fig. 20.—Equivalent circuit of photo-voltaic cell.

in the external load is given as a fraction of the primary photo-current, I_p, by

$$I = \frac{I_p R_i}{R_s + R_i + R_L}$$

The frequency response of the actual photo-voltaic effect is practically instantaneous, but the response of the cell falls off steeply above a few thousand cycles per second, owing to the high capacitance of the cell ($\sim 0{\cdot}5\,\mu$F), which is, in effect, in parallel with the external load resistance.

The open-circuit voltage generated rises rapidly at first with illumination (i.e. light flux per unit area of the cell) up to some 10 or 20 foot-candles, and thereafter more gradually, reaching a maximum of about 0·5 volt at a very high illumination. The only current flow in the cell under these conditions is the reverse current through the internal resistance, and since this resistance decreases with increasing illumination, it follows that the curve of the open-circuit voltage plotted against illumination will flatten at higher values.

If a load resistance is coupled to the cell, the current flowing through it will be closely proportional to the total light flux falling on the cell if the resistance is low compared with the internal resistance. However, as the load resistance is increased the current flowing through it falls from proportionality progressively more and more at higher values of light flux.

The selenium barrier-layer cell exhibits a relatively small and temporary fatigue, from 1 to 5%, increasing with the load resistance. However, the long-term stability, or life, of such cells is very good.

The sensitivity attainable is 100–500 μA/lumen or—in terms of power conversion—50 μW/lumen. The cell is usually housed in a Bakelite case and the sensitive surface is protected by a glass, thus making a robust and simple device. These features, combined with the fact that a power source is not needed, make such cells very suitable for use as light meters.

(12) BIBLIOGRAPHY

(1) Zworykin, V. K., and Ramberg, E. G.: 'Photo-Electricity and its Applications' (Wiley, New York, 1949).
Sommer, A.: 'Photoelectric Tubes' (Methuen, London, 1953).

(2) Becquerel, E.: 'Studies of the Effect of Actinic Radiation of Sunlight by means of Electric Currents', *Comptes Rendus*, 1839, **9**, p. 145.

(3) Smith, W.: 'The Action of Light on Selenium', *Journal of the Society of Telegraph Engineers*, 1873, **2**, p. 31. See also: *American Journal of Science*, 1873, **5**, p. 301.

(4) Hertz, H.: 'On Very Rapid Electrical Oscillations', *Annalen der Physik*, 1887, **31**, p. 421.

(5) Thomson, J. J.: 'On the Masses of the Ions in Gases at Low Pressures', *Philosophical Magazine*, 1899, **48**, p. 547.

(6) Einstein, A.: 'A Heuristic Standpoint concerning the Production and Transformation of Light', *Annalen der Physik*, 1905, **17**, p. 132.

(7) Planck, M.: 'On the Theory of the Law of Energy Distribution in the Normal Spectrum', *Verhandlungen der Deutschen Physikalische Gesellschaft*, 1900, **2**, p. 237.

(8) Taylor, J. B., and Langmuir, I.: 'The Evaporation of Atoms, Ions and Electrons from Caesium Films on Tungsten', *Physical Review*, 1933, **44**, p. 423.

(9) Becker, J. A.: 'Thermionic and Absorption Characteristics of Caesium on Tungsten and Oxidized Tungsten', *ibid.*, 1926, **28**, p. 341.

(10) Kingdon, K. H.: 'Electron Emission from Adsorbed Films on Tungsten', *ibid.*, 1924, **24**, p. 510.

(11) Koller, L. R.: 'Some Characteristics of Photo-Electric Tubes', *Journal of the Optical Society of America*, 1929, **19**, p. 135.

(12) Asao, S., and Suzuki, M.: 'Improvement of Thin-Film Caesium Photo-Electric Tube', *Proceedings of the Physical-Mathematical Society of Japan*, 1930, **12**, p. 247.

(13) McGee, J. D., and Lubszynski, H. G.: 'E.M.I. Cathode-Ray Television Transmission Tubes', *Journal I.E.E.*, 1939, **84**, p. 468.

(14) Goerlich, P.: 'On Composite Transparent Photo-Cathodes', *Zeitschrift für Physik*, 1936, **101**, p. 335.

(15) Sommer, A.: 'New Photoemissive Cathodes of High Sensitivity', *Review of Scientific Instruments*, 1955, **26**, p. 725.

(16) Ives, H. E., Olpin, A. R., and Johnsrud, A. L.: 'The Distribution in Direction of Photo-Electrons from Alkali Metal Surfaces', *Physical Review*, 1928, **32**, p. 57.

(17) Brady, J. J.: 'Energy Distribution of Photo-Electrons as a Function of the Thickness of a Potassium Film', *ibid.*, 1934, **46**, p. 768.

(18) McGee, J. D.: British Patent No. 443777, 1934.

(19) Rajchman, J. A.: 'Le courant résiduel dans les multiplicateurs d'éléctrons éléctrostatiques' (Thesis, Kundig, Geneva, 1938).

(20) Sommer, A., and Turk, W.: 'New Multiplier Phototubes of High Sensitivity', *Journal of Scientific Instruments*, 1950, **27**, p. 113.

(21) McGee, J. D.: 'Television Pick-Up Tubes', *Journal I.E.E.*, 1955, **1** (New Series), pp. 404 and 502.

(22) *Proceedings of the Institute of Radio Engineers*, 1955, **43**, No. 12, Papers Nos. 5578–5582.

(23) Case, T. W.: 'Thalophide Cell—a New Photo-Electric Substance', *Physical Review*, 1920, **15**, p. 289.

(24) Von Hippel, A., and Rittner, E. S.: 'Thallous Sulphide Photoconductive Cells, II—Theoretical Discussion', *Journal of Chemical Physics*, 1946, **14**, p. 370.

(25) Elliott, A.: 'Recent Advances in Photocells for the Infra-Red', Chapter III of 'Electronics', edited by B. Lovell (Pilot, London, 1947).

(26) Grondahl, L. O.: 'The Copper-Cuprous Oxide Rectifier and Photoelectric Cell', *Reviews of Modern Physics*, 1933, **5**, p. 141.

(27) Lange, B.: 'New Kind of Photo-Electric Cell', *Physikalische Zeitschrift*, 1930, **31**, pp. 139 and 964.

(28) Schottky, W.: 'Origin of Photo-Electrons in Copper-Oxide Photo-Electric Cells', *ibid.*, p. 913.

(29) Duhme, E., and Schottky, W.: 'Rectifying and Photo-Electric Effects at Contacts between Cuprous Oxide and Sputtered Electrodes', *Naturwissenschaften*, 1930, **18**, p. 735.

(30) Mott, N. F.: 'Note on Copper Cuprous-Oxide Photocells', *Proceedings of the Royal Society*, A, 1939, **171**, p. 283.

(31) Waddell, J. M., Mayer, S. E., and Kaye, S.: 'A Germanium Diffused-Junction Photo-Electric Cell', *Proceedings I.E.E.*, Paper No. 1860 R, April, 1955 (**102** B, p. 757).

621.316.825:621.317.794:621.317.38.029.64

Paper No. 2381 R
Sept. 1957

MEASUREMENTS OF EFFICIENCY OF BOLOMETER AND THERMISTOR MOUNTS BY IMPEDANCE METHODS

By J. A. LANE, M.Sc., Associate Member.

(*The paper was first received* 22*nd February, and in revised form* 11*th April,* 1957.)

SUMMARY

Methods requiring only measurements of impedance are used to investigate the efficiency of resistance-type milliwattmeters at frequencies of 9·2 Gc/s and 3 Gc/s (wavelengths of 3·26 cm and 10·0 cm). For fine-wire bolometers the results agree with those obtained by calorimetric methods and are consistent with the few results previously available at lower frequencies. Accurate values of efficiency cannot be obtained by impedance methods, in their present form, for thermistors at microwave frequencies owing to the uncertainty regarding the precise impedance of the semi-conducting bead.

(1) INTRODUCTION

Resistance-type milliwattmeters such as thermistors and bolometers are widely used for the measurement of power at microwave frequencies, and techniques for the calibration of these devices are of general interest in many applications. As an alternative to the somewhat difficult procedure requiring a comparison against standard equipment operating at a level of several watts, other methods have been suggested in which the efficiency of bolometer and thermistor mounts is calculated from measurements of impedance.[1,2] It has been claimed, for example, that an error limit of no more than $\pm 1{\cdot}5\%$ can be achieved in the calibration of bolometer mounts at frequencies of 0·6–3·0 Gc/s. This degree of accuracy certainly compares favourably with that obtainable in comparison methods. At the present time, however, the validity of this novel approach to the problem of milliwattmeter calibrations remains unconfirmed. In particular, the accuracy of the results obtained with thermistor mounts is subject to some doubt, especially at frequencies above 3 Gc/s.

It is the purpose of the paper to summarize typical results of milliwattmeter efficiency obtained by impedance methods at frequencies of 3 Gc/s and 10 Gc/s. The accuracy of the results is examined by comparing them with similar data obtained from comparison measurements using standard power-measuring equipment.

(2) BASIS OF METHODS

The development of the analysis used in this investigation is largely due to workers at the National Bureau of Standards, and the theoretical procedure followed by them is given in detail in their original papers.[1,2] For the present purpose only an outline of the methods, with the relevant equations, need be given.

The properties of any waveguide structure may be analysed by two general methods. It is possible, for example, to derive a series of network equations which are similar in form to those used in analytical methods at low frequencies: in this method, the equations are frequently summarized by the corresponding impedance (or admittance) matrix. Alternatively, the fields in the waveguide may be described in terms of the amplitudes of the incident and reflected, or scattered, waves. This approach yields an array of scattering coefficients (a scattering matrix). The two matrices are related in that one can be calculated if the other is known. Both these techniques have been employed in the present problem, and in each case the bolometer or thermistor mount is regarded as a linear passive network connecting the input reference terminals (which may be located at any suitable plane in the input waveguide or coaxial line) and the resistive bead or wire. As will be evident from the results given below, a difficulty common to both methods is the satisfactory representation, in terms of an impedance or a scattering coefficient, of the load at the output terminals.

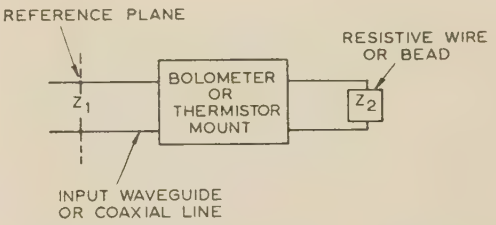


Fig. 1.—Bolometer or thermistor mount as a two-terminal network.

Consider the mount assembly shown in Fig. 1, in which Z_1 is the normalized input impedance at a reference plane in the input waveguide, and Z_2 is the impedance of the bead or wire in the resistive load. Kerns[1] has derived equations for the efficiency of power transfer, η, of such a network in terms of three measured values of Z_1 (say Z_1', Z_1'', Z_1''') corresponding to three values of Z_2 (say Z_2', Z_2'', Z_2'''). Suppose, for example, that the mount is made reflectionless with a termination Z_2''. Then $Z_1'' = 1$. If Z_2' and Z_2''' are expressed by values relative to Z_2'', we have

$$\eta = \left| \frac{(1 - Z_1')(1 - Z_1''')(Z_2''' - Z_2')}{(1 - Z_2')(1 - Z_2''')(Z_1''' - Z_1')} \right| \quad . \quad . \quad (1)$$

In the modified treatment developed by Beatty and Reggia,[2] the scattering equations of a 2-terminal network are solved to give an equivalent expression for η in terms of the reflection coefficient at the input and output terminals. Consider a loosely-coupled probe located in the input transmission line following a matched generator, and let the voltages induced in the probe be V_1, V_2 and V_3 for values of resistive load, R, of R_1, R_2 and R_3. If the mount is tuned so that the input reflection coefficient is zero when $R = R_2$, say, and the probe is located at a fixed position such that its response is a maximum when $R = R_1$, we have

$$\eta = k(V_1 - V_2)(V_2 - V_3)/(V_1 - V_3)V_2 \quad . \quad . \quad (2)$$

where $$k = 2R_2(R_3 - R_1)/(R_2 - R_1)(R_3 - R_2)$$

and $$R_1 < R_2 < R_3$$

If, in addition, the mount has a relatively high efficiency, say 80% or greater, it can be shown that the input reflection coefficients, ρ_1, ρ_3, corresponding to resistive loads, R_1, R_3, differ in phase by almost exactly 180°. The position of a voltage

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper is an official communication from the Radio Research Station, Department of Scientific and Industrial Research.

minimum in the input waveguide when $R = R_1$ is therefore almost identical with that of a voltage maximum when $R = R_3$. Under these conditions it is possible to represent the equivalent network in terms of real parameters. (If necessary, a small correction can be made, as described by Beatty and Reggia, to allow for the slight departure from co-linearity in the vectors representing ρ_1 and ρ_3.)

Now the determination of mount efficiency is essentially a measurement of attenuation, and it is possible to derive a third expression for η which is similar to that frequently used in measuring the attenuation coefficient of transmission lines. Suppose that the mount assembly is matched (i.e. $\rho = 0$) for a given value of terminating resistance, R. Let the modulus of the input reflection coefficient, $|\rho|$, be $|\rho_1|$ with a different value of R, say R_1. The reflection coefficient, $|\rho'|$, at the output terminals is given by

$$|\rho'| = (1 - R_1/R)/(1 + R_1/R) \quad . \quad . \quad . \quad . \quad (3)$$

The attenuation coefficient, α, is then given, as in transmission line theory, by

$$|\rho'| = |\rho_1|\varepsilon^{2\alpha} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The attenuation between input and output terminals is thus given by $8{\cdot}69\alpha$ dB. (Strictly speaking, the mount attenuation is required with the network terminated in a load R rather than R_1; however, the results given below suggest that no appreciable error is introduced by the procedure outlined above.)

Kerns[1] has suggested that the radio-frequency impedance of typical bolometer wires and thermistor beads is approximately equal to the d.c. resistance, at any rate at frequencies up to a few gigacycles per second. This assumption is probably valid for the small fine-wire bolometers now available; consequently, eqns. (1), (2) and (4) should give valid results for η for this type of milliwattmeter. The thermistor, on the other hand, is known to have a finite self-capacitance which renders the impedance frequency-dependent above about 3 Gc/s. Accurate results are not to be expected, therefore, with thermistor mounts in this region of the spectrum.

(3) EXPERIMENTAL RESULTS AND DISCUSSION

The experiments consisted simply of measurements of the input voltage-standing-wave ratio (v.s.w.r.) for several bolometer and thermistor mounts with the resistive load operated at various known d.c. resistances. Typical results for a thin-wire bolometer at a frequency of 9·2 Gc/s (wavelength 3·26 cm) are given in Table 1, together with the corresponding results obtained in calibration measurements using a water calorimeter.

Table 1

EFFICIENCY OF BOLOMETER MOUNT, $f = 9{\cdot}2$ Gc/s

Efficiency, η			
Impedance methods			Calorimetric method
Eqn. (1)	Eqn. (2)	Eqn. (4)	
0·90	0·92	0·90	0·89

The values of η quoted in Table 1 all have an error limit of the order of $\pm 3\%$. There appears to be no significant difference, therefore, between the results obtained in the impedance methods and the calorimetric method for determining bolometer-mount efficiency, even at a frequency as high as 9·2 Gc/s. Provided, therefore, that the measurements of v.s.w.r. can be made with a high order of accuracy, the techniques outlined above afford a useful alternative to calorimetric methods for the calibration of bolometer mounts in the frequency band 0·5–10 Gc/s. The values of η given by Beatty and Reggia[2] for coaxial bolometer mounts in the band 0·6–3·0 Gc/s are in the range 0·96–0·99. These results are consistent with those given in Table 1, for a general decrease in η with increasing frequency is to be expected.

Measurements were also made on several thermistor mounts, with less satisfactory results. At a frequency of 9·2 Gc/s, the impedance methods gave values of η in the range 0·68–0·77, compared with 0·87 as determined by calorimetric techniques. At a frequency of 3 Gc/s, the latter gave $\eta = 0{\cdot}97$ for a waveguide thermistor mount, compared with a value of 0·87 determined by impedance methods. For this type of milliwattmeter, therefore, the impedance methods underestimate the mount efficiency.

Typical thermistor beads have a self-capacitance which is thought to be[3] of the order of 0·2 pF. If this is so, the assumption suggested by Kerns[1] of a resistive load for this type of milliwattmeter is obviously invalid, at any rate for frequencies in the microwave band. Although it is possible, with certain limitations, to calculate efficiency values in Kern's analysis for loads having a reactive component, this procedure requires a precise knowledge of the load impedance. In any case, an equivalent circuit for a thermistor bead consisting of the d.c. resistance shunted by the self-capacitance is probably not a complete explanation. It is possible to show that, with such a circuit, the measured efficiency should vary with bias current, at frequencies of the order of 10 Gc/s and above, owing to the varying proportion of the input power which is dissipated in the tie-wires supporting the thermistor bead. However, the observed variation at 30 Gc/s is much smaller than the theoretical variation,[3] and at 10 Gc/s it appears to be negligible.[4] It seems probable, therefore, that the effective resistance of the semiconducting bead at high frequencies varies less rapidly with temperature than the d.c. resistance, as suggested by Collard, Nicoll and Lines.[3]

The results of further impedance measurements on thermistors, in both coaxial lines and waveguides over a wide frequency range, were analysed using several equivalent circuits, but the results were generally inconsistent with any simple model.

(4) ACKNOWLEDGMENTS

The work was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

(5) REFERENCES

(1) KERNS, D. M.: 'Determination of Efficiency of Microwave Bolometer Mounts from Impedance Data', *Journal of Research, National Bureau of Standards*, 1949, **42**, p. 579.

(2) BEATTY, R. W., and REGGIA, F.: 'An Improved Method of Measuring Efficiencies of Ultra-High Frequency and Microwave Bolometer Mounts', *ibid.*, 1955, **54**, p. 321.

(3) COLLARD, J., NICOLL, G. R., and LINES, A. W.: 'Discrepancies in the Measurement of Microwave Power at Wavelengths below 3 cm', *Proceedings of the Physical Society*, 1950, **63**, p. 215.

(4) LANE, J. A.: 'The Measurement of Power at a Wavelength of 3 cm by Thermistors and Bolometers', *Proceedings I.E.E.*, Paper No. 1918 R, November, 1955 (**102** B, p. 819).

621.385.1.032.29

Paper No. 2384 R
Sept. 1957

GROWTH OF ANODE-TO-GRID CAPACITANCE IN LOW-VOLTAGE RECEIVING VALVES

By F. H. REYNOLDS, B.Sc.(Eng.), Associate Member, C. B. JOHNSON and M. W. ROGERS.

(*The paper was first received 25th January, and in revised form 24th April*, 1957.)

SUMMARY

The capacitance between the anode and the control grid of a certain type of receiving pentode has been found to increase with life, the rate of growth being dependent on the operating conditions of the valve and on the material and processing schedule of the anode. It is shown that the phenomenon is due to the transfer of impurity carbon from the anode to the mica insulators.

(1) INTRODUCTION

The receiving-type thermionic valve is now used in many fields outside its traditional place in a radio receiver. It is usually found that in newer applications a more exacting performance is required from the valve, at least in respect of some of its properties; there is perhaps no better example of this than the stringent life specification placed on the amplifier valves for submerged telephone repeaters.

As a result of this wider use, a concerted attack has recently been made on some of the outstanding valve problems and, to quote the case of submerged-repeater valves again, life-stability has received extensive treatment. The point has now been reached where the common electrical forms of valve failure during life—loss of cathode emission and growth of core/coating interface resistance—can be controlled to an appreciable extent. Unfortunately even a valve free of these troubles may still develop other important defects, of which the growth of inter-electrode capacitances and the generation of excessive noise can be serious enough to rank as causes of failure. The noise problem in Post Office submerged-repeater valves has already been discussed;[1] the growth of capacitance between the anode and the control grid has been sufficiently troublesome to warrant the attention described here.

The growth of anode-to-grid capacitance in valves has been observed by other workers, notably James and Humphreys.[2] Usually, the phenomenon is attributed to the presence of thin films on the valve insulators produced by cathode evaporation, but there are probably several other processes which can also engender capacitance growth. In this work, only one such process is isolated and given detailed attention. The investigations have been carried out exclusively on valves having cathode cores of pure platinum, since this is the standard core material for Post Office submerged-repeater valves. It will be shown, however, that the use of platinum has been advantageous in isolating the phenomenon without being necessary for its occurrence. There is consequently no reason why the processes described should not occur in many commercial valves having cathode cores of other materials.

(2) THE VALVE EMPLOYED FOR THE INVESTIGATIONS

The Post Office valve type 6P12 is a modified version of the well-known type CV138, with the active electrode structure redesigned for operation at lower h.t. voltages (screen grid at 60 volts). The electrode assembly is mounted on a conventional pinch instead of a pressed-glass base, and the control-grid connection is brought out through the dome of the bulb.

The term 'standard valve' will be applied to a valve which has a cathode core of thermo-pure platinum with a double-carbonate coating, grid support rods of nickel, grid helix wires of molybdenum and an anode of a gas-free grade of nickel. The highest-quality ruby mica has been used for the insulators, which are, moreover, uncoated.

The capacitance between the anode and the control grid of a new valve, measured 'cold' with all the other electrodes and the bulb metallizing earthed, is approximately 0·013 pF.

(3) CHARACTERISTICS OF THE CAPACITANCE-GROWTH PHENOMENON

It was discovered at an early stage that the rate of growth of anode-to-grid capacitance in the type 6P12 valve varied widely from sample to sample, some valves not exhibiting the effect at all. Part of the variation may be attributed to different operating conditions, a study of which throws much light on the problem, as will be discussed later. A wide spread of behaviour was, however, observed between standard valves operated under the same conditions, as is shown by Fig. 1, for which all the measurements were made at 1 kc/s. An initial rise of capacitance is often detectable after only a few hundred hours' operation, the value then increasing until a ceiling of about 0·090 pF is reached after perhaps 1 000–2 000 hours. In other valves the anode-to-grid capacitance may still be rising steadily after 16 000 hours. Values as high as 0·2 pF are sometimes obtained, but these tend to be unstable and susceptible to change on subjecting the valve to mechanical shock.

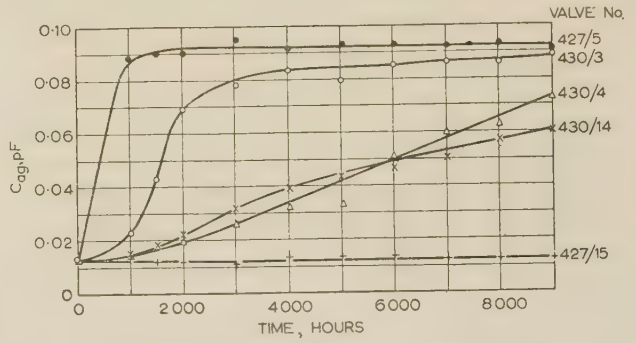


Fig. 1.—Growth of anode-to-grid capacitance: typical curves for individual valves.

An enhanced anode-to-grid capacitance is always accompanied by a shunt resistance component.

If air is admitted to a valve exhibiting an enhanced anode-to-grid capacitance near the ceiling value, the capacitance remains substantially unaffected. Very small enhanced capacitances tend to disappear on exposure.

(3.1) Frequency Dependence

The dependence of the observed anode-to-grid capacitance on

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The authors are at the Post Office Research Station.

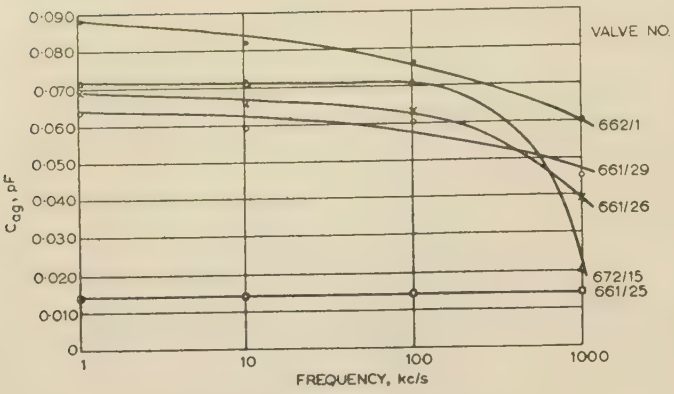


Fig. 2.—Variation of anode-to-grid capacitance with frequency.

the frequency of measurement over the range 1–1 000 kc/s is shown by Fig. 2, which is plotted from measurements on valves after 8000 hours of life. The graphs show that an enhanced value falls as the frequency rises, particularly over the range 100–1 000 kc/s, but may still be four times the zero-time value of anode-to-grid capacitance at 500 kc/s or even at 1 Mc/s.

(4) LIFE-TEST RESULTS

(4.1) Routine Life Tests

All type-6P12 valves satisfying an acceptance specification at the time of manufacture are, as a routine procedure, subjected to a period of life test under conditions which simulate operation in a repeater amplifier. After not less than 4000 hours' operation some valves are selected for repeater use, while the remainder continue on test as 'shadow' batches. It is from these latter valves that the information in this Section is drawn.

(4.1.1) Tests of Standard Valves.

Two standard life-test conditions are employed, one to simulate use as the output valve of an amplifier and the other to simulate voltage-amplifier conditions. The conditions are:

Output: $V_a = 90$ volts, $V_{g2} = 60$ volts, $I_a = 6$ mA
Input: $V_a = 40$ volts, $V_{g2} = 60$ volts, $I_a = 4$ mA

where V_a, V_{g2} and I_a are the anode voltage, screen-grid voltage and anode current respectively.

The growth of anode-to-grid capacitance, C_{ag}, in two large batches of standard valves, one operated under input conditions and the other under output conditions, is shown by Table 1.

Table 1

INFLUENCE OF OPERATING CONDITIONS

Length of test	Output valves (600 valves)		Input valves (500 valves)
	$C_{ag} > 0{\cdot}020$ pF	$C_{ag} > 0{\cdot}050$ pF	$C_{ag} > 0{\cdot}020$ pF
hours	%	%	%
1000	37	25	0
2000	52	40	0
3000	58	45	0

Evidently there is a fair probability that an output valve will exhibit an enhanced anode-to-grid capacitance, but the input valves are completely free of the phenomenon after 3000 hours' operation. Later measurements have shown no sign of capacitance growth after over 24 000 hours under input conditions.

(4.1.2) Tests of Modified Valves.

In another series of life tests, valves have been used which differ only from the standard valves in that heat-radiating fins have been fastened to the backs of the anodes. In Table 2 the

Table 2

EFFECT OF ANODE COOLING FINS

Length of test	Standard valves (100 valves)		Finned valves (100 valves)	
	$C_{ag} > 0{\cdot}020$ pF	$C_{ag} > 0{\cdot}050$ pF	$C_{ag} > 0{\cdot}020$ pF	$C_{ag} > 0{\cdot}050$ pF
hours	%	%	%	%
3000	65	51	0	0
8000	68	55	51	18

growths of anode-to-grid capacitance in batches of finned and unfinned (standard) valves operated under output conditions are compared.

Evidently the addition of fins checks the rate of growth of anode-to-grid capacitance but does not prevent it.

(4.1.3) Discussion.

From the results of the routine life tests it is already possible to infer that the anode of the valve plays some important part in the mechanism of capacitance growth. The anode temperature of the standard valve operated under output conditions is 350° C, compared with 250° C for the same valve under input conditions, but the currents and voltages are also different for the two batches. With the finned and standard valves, however, the electronic conditions are the same for both batches, and yet the growth of capacitance is delayed by the fins, which reduce the anode temperature from 350 to 285° C. The tentative conclusion may therefore be drawn that the higher the anode temperature the more rapid is the growth of anode-to-grid capacitance.

(4.2) Special Life Tests

As a result of the above conclusions, the part played by the anode has received further attention, although the possibility that the cathode can also influence the rate of growth has not been overlooked. The valves for the special life tests were all of normal repeater quality, but changes were introduced as detailed below. Standard output life-test conditions were used unless otherwise indicated.

(4.2.1) Influence of Anode Material.

Batches of otherwise standard valves have been made with different anode materials, all anodes being processed in the same way. The valves were operated under output conditions, a check being made that their anode temperatures approximated to 350° C. The results are shown in Table 3.

Clearly, the commercial nickels have some property, not possessed by the other metals investigated, which is necessary for growth of anode-to-grid capacitance.

After an appreciation of the sensitivity of the capacitance-growth phenomenon to the material of the anode, an attempt was made to explain the spread of behaviour of seemingly similar valves (as in Section 3) by the variation in impurity content of nominally similar anodes. Accordingly, the life-test results for output valves were re-examined for any relationship with the supply batch of nickel from which the anodes had been made. A correlation was obtained which is shown in Table 4. In this Table, all the valves happen to be fitted with anode cooling fins, but this does not invalidate the comparison.

Table 3

INFLUENCE OF ANODE MATERIAL

Anode material	Number of valves in batch	Length of test	Fraction with $C_{ag} > 0{\cdot}020$ pF
		hours	%
Commercial gas-free nickel (standard)	100	24 000	72
Commercial cathode nickel ..	6	20 000	16
Spectrographically pure nickel	10	20 000	0
Spectrographically pure nickel + 0·07% magnesium	10	3 000	0
Molybdenum	30	28 000	0
Platinum	30	28 000	0

Table 4

INFLUENCE OF ANODE IMPURITIES

Length of test	Fraction of batch with $C_{ag} > 0{\cdot}020$ pF			
	Batch A (100 valves)	Batch B (59 valves)	Batch C (20 valves)	Batch D (20 valves)
hours	%	%	%	%
1 000	0	14	50	90
2 000	0	65	95	100
3 000	0	70	95	100
6 000	36	81	100	100
8 000	51	Off test	Off test	Off test

Between the batches A–D there are appreciable variations in the impurity levels of several elements. In the case of carbon, the level increases with the sequence A to D, but no similar trend can be observed with any other element.

(4.2.2) **Influence of Anode Processing.**

In the tests previously described the anodes were processed by firing *in vacuo* for a total of six minutes at 1 000° C, after which they were stored *in vacuo* until just prior to assembly of the valves.

The effect of substituting a period of hydrogen firing for the above schedule is shown in Table 6. One batch had the normal anode processing; for the other, the anodes were fired for 3 hours in dry hydrogen (dew point −60° C) at 1 150° C. All the valves in Table 6 had anodes of batch-B nickel. The anodes which received the normal vacuum treatment were also finned, whereas the hydrogen-fired anodes were of standard design. However, since the absence of fins is conducive to capacitance growth (by the results given in Section 4.1.2) it is permissible to conclude that the complete absence of capacitance growth in the valves with hydrogen-fired anodes is due to the hydrogen treatment.

Analysis of the impurity content of the anode nickel after the standard vacuum treatment shows no measurable change in the impurity content. After the hydrogen-firing treatment, the impurity concentrations are again unaltered, with the exception of magnesium and carbon, for which typical results are given in Table 7.

The long period of hydrogen firing is seen to reduce the magnesium concentration to less than half and the carbon content to one-seventh of the original values.

Table 5

ANALYSIS OF ANODE NICKELS

Batch	Impurity content											
	Carbon	Silver	Alumi-nium	Cobalt	Chromium	Copper	Iron	Magnesium	Manga-nese	Lead	Antimony	Silicon
	%	%	%	%	%	%	%	%	%	%	%	%
A	0·009	0·000 1	0·000 5	<0·05	0·005	0·06	0·03	0·11	0·005	0·005	0·02	0·03
B	0·028	0·000 1	0·000 5	0·05	<0·005	0·005	0·03	0·06	0·05	0·005	0·02	0·02
C	0·032	0·000 2	0·001	0·05	0·005	0·01	0·02	<0·05	0·005	0·005	0·01	0·02
D	0·041	<0·000 1	0·001	<0·05	0	0·005	0·02	0·05	0·005	0·005	0·01	0·02
Cathode grade ..	0·006	<0·000 1	0·001	0·05	0·001	0·005	0·02	0·025	0·001	0·005	<0·001	0·01

The rate of growth of anode-to-grid capacitance is seen to increase, according to the supply batch used, in the sequence A, B, C, D. Analyses of the impurity contents of these supply batches are given in Table 5, the last line giving an analysis of the commercial grade of cathode nickel used for the results in Table 3.

Table 6

EFFECT OF HYDROGEN FIRING OF ANODES

Length of test	Fraction with $C_{ag} > 0{\cdot}020$ pF	
	Vacuum-fired anodes (59 valves)	Hydrogen-fired anodes (56 valves)
hours	%	%
1 000	14	0
2 000	65	0
3 000	70	0

Table 7

EFFECT OF HYDROGEN-FIRING ON NICKEL IMPURITIES

Treatment (of batch-B nickel)	Magnesium content	Carbon content
	%	%
Raw material (or after standard vacuum treatment)	0·06	0·028
After 3 hours at 1 150° C in dry hydrogen	0·025	0·004

(4.2.3) **Effect of Varying Cathode Temperature.**

Batches of standard valves have been tested under output conditions with different cathode temperatures, values of 4·0 (900° K), 4·5 (950° K), 5·0 (990° K) and 6·0 volts (1 060° K) being selected for the heater voltage (cathode core temperature in brackets) instead of the standard value of 5·5 volts (1 020° K).

The (unfinned) anodes of these valves were made from batch-A

Table 8

INFLUENCE OF CATHODE TEMPERATURE

Length of test	Fraction of group with $C_{ag} > 0{\cdot}020$ pF after operation at heater voltage V_h (approximately 10 valves per batch)				
	$V_h = 4{\cdot}0$ volts	$V_h = 4{\cdot}5$ volts	$V_h = 5{\cdot}0$ volts	$V_h = 5{\cdot}5$ volts	$V_h = 6{\cdot}0$ volts
hours	%	%	%	%	%
1 000	12·5	40	11	3·5	33
2 000	38	60	22	37	53
3 000	50	70	33	—	58

material and received the normal vacuum processing treatment. The results, given in Table 8, show a considerable spread of behaviour between the batches, but there seems to be no possibility of attributing the spread to the variation in cathode temperature.

(4.2.4) **Discussion.**

The results of the special life tests confirm the important role of the anode and suggest that the part played by the cathode is small; at least the cathode does not exert a rate-determining influence.

Of the anode materials listed in Table 3, only the use of commercial grades of nickel resulted in capacitance growth (within 20 000 hours). The addition, to an otherwise spectrographically pure nickel, of magnesium to give a concentration typical of a commercial nickel does not yield an anode material which is likely to engender capacitance growth. Comparing batches A–D of commercial nickel, it appears that the faster the rate of capacitance growth the higher is the level of the carbon impurity in the anode. This correlation is confirmed by the effect of hydrogen firing, which apparently eliminates capacitance growth and leaves most of the impurity concentrations unaltered except for the carbon level, which is drastically reduced.

The amount of carbon in an anode of nickel emerges as the crucial factor in this phenomenon of growth of anode-to-grid capacitance. Under output conditions, a level of 0·025% or more will produce capacitance growth in a very short time, even if cooling fins are fitted. Even 0·009% will produce fairly rapid growth, although the onset can be delayed for perhaps 5 000 hours by fitting fins. At 0·006% (the concentration in the anodes made of cathode-grade nickel) the effect has only been positively identified on one valve and that was after 20 000 hours' operation, while growth has never yet been observed at the 0·004 % level obtained by hydrogen stoving.

(5) THIN FILMS AS THE CAUSE OF CAPACITANCE GROWTH

As is shown by the work of Howe[3,4] and James and Humphreys,[2] a capacitance which falls as the frequency of measurement rises is characteristic of the self-capacitance of a very high resistance such as is obtained from thin conducting films. This capacitance will also be shunted by a resistance. This information, coupled with the difficulty of conceiving an alternative explanation, led to the assumption at the beginning of these investigations that the growth of capacitance was due to the formation, somewhere, of a thin conducting film.

There are two possible locations for this film (Fig. 3): (*a*) on the glass-bulb interior, providing a capacitive coupling between the control-grid lead and the anode; and (*b*) on the mica insulators, linking the regions where the anode and grid are located. Both possibilities were explored by the deliberate deposition of films, but only the location (*b*) simulated the enhanced capacitance of a life-tested valve. An increased capacitance can be produced by mechanism (*a*), but it is entirely cancelled out upon coating the bulb with its usual earthed metallizing. It was found also that it is not necessary for the film on the insulators to make direct contact with the grid and the anode. Since these electrodes, and particularly the anode, are not tightly located in the mica insulators, it is probable that direct contact between electrode and the film will normally not occur. No matter how much the resistance of the film may fall, therefore, the increase in anode-to-grid capacitance cannot exceed that of the two capacitances, anode-to-film and grid-to-film in series, thus explaining why a ceiling value of anode-to-grid capacitance is usually obtained. In the unusual case where direct contact between the electrode and the film does occur, the normal ceiling value would be exceeded. It is, moreover, reasonable that such an enhanced value should vary when the valve is subjected to mechanical shock.

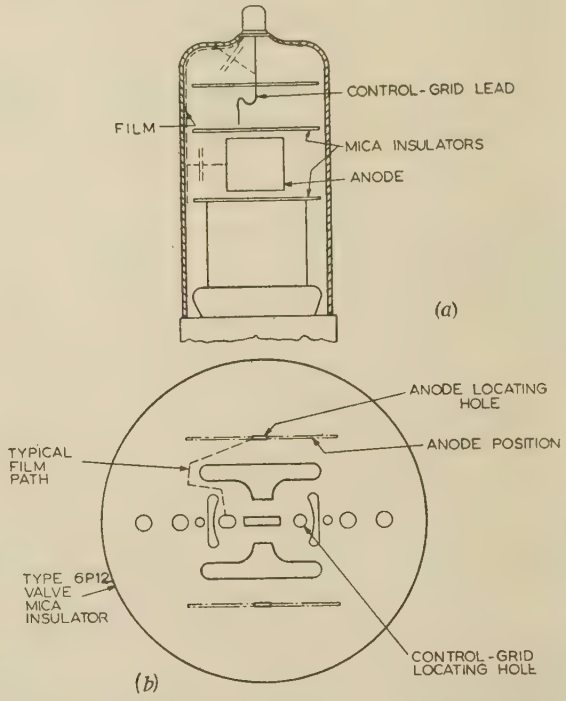


Fig. 3.—Location of film for growth of anode-to-grid capacitance.

(5.1) Detection of Thin Films

In the attempt to locate positively the presence of a conducting film on the mica insulators, the supposition was made initially that, if the insulators received material by evaporation from any part of the valve, so also should the glass bulb. A film deposited on the inner wall of a glass bulb can, however, be detected by means of the triple-band method recently described.[5] A brief summary of this method is given below.

(5.1.1) **The Triple-Band Method.**

Three equispaced conducting bands of similar width are formed on the outside of the valve bulb over the region where it is desired to detect an internal conducting film. The detection of a film is then achieved with the aid of 3-terminal capacitance measurements, the middle band being connected to the earth terminal of the bridge and the outer bands to the line terminals. In the absence of a conducting film, a low capacitance is measured, since the two line bands are effectively screened by the earthed band. As soon as a conducting film is deposited on

the inner wall of the bulb opposite the triple-banded region, the earthed band is by-passed and the bridge measures an increased capacitance which depends on the conductivity of the deposited film.

As an extension of the method, the two line bands may be subdivided into a number of segments and the capacitance between each pair of segments on either side of the earthed band measured. The film thickness can then be plotted round the circumference of the bulb.

Triple-band measurements made on the bulbs of standard type-6P12 valves operated under output conditions showed that increase in anode-to-grid capacitance is subsequently followed by a rise in triple-band capacitance. If the line bands in such a valve are then subdivided, the triple-band capacitance may be plotted round the bulb. This is shown by Fig. 4, where the

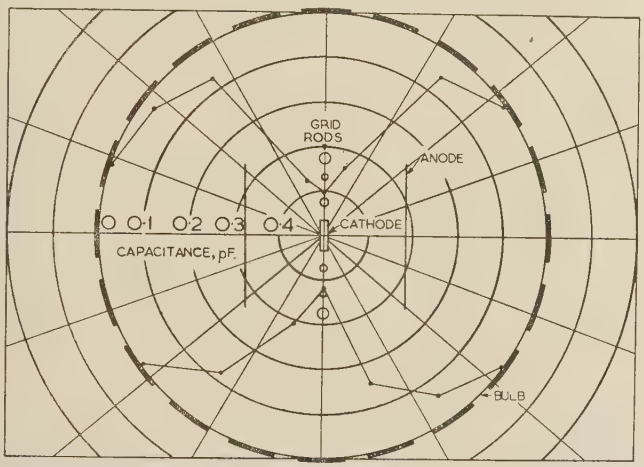
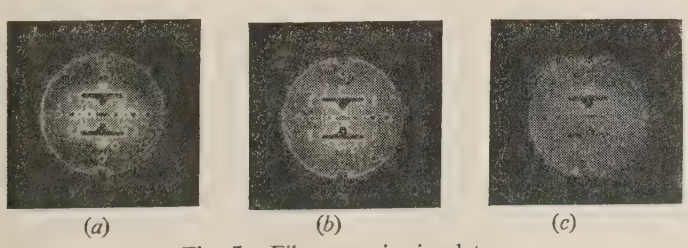


Fig. 4.—Distribution of conducting film around inner surface of bulb of valve exhibiting growth of anode-to-grid capacitance.

capacitances have been plotted radially inwards using the circle representing the bulb as datum. It is seen at once that the conducting material must have emanated from the region of the electrode structure, and, moreover, the immediate source cannot be the cathode, otherwise the curve would show depressions along the line of the grid support-rods. Further study of this diagram yields the conclusion that the film material must have emanated from the inner surfaces of the anodes.

(5.1.2) Visual Search for Deposited Film.

When the results given in the previous Section became available, the efforts to locate a film on the mica insulators were redoubled. Eventually, using the insulators taken from valves which had operated under predisposing conditions for nearly 10 000 hours and after cleaving the mica to about 0·002 in thickness, a faint black deposit was detected. Such a deposit has been used to make the photographic print in Fig. 5(*a*); it is seen to be immediately in front of the anodes, with a clear line of demarcation under the anode edge (the position of the anode is marked on Fig. 3). For comparison, Fig. 5(*b*) shows the pattern which would have been obtained if the film material had evaporated from the surface of the cathode. Subsequently, using batch-B vacuum-processed anode nickel and omitting cooling fins, detectable black deposits have been obtained after only 1 500 hours of operation.



(*a*) (*b*) (*c*)

Fig. 5.—Films on mica insulators.

(*a*) Deposit on insulator taken from valve after 10 000 hours of operation.
(*b*) Film pattern obtained from cathode evaporation.
(*c*) Deposit on insulator taken from special valve having one anode plate of nickel and the other of platinum after 2 500 hours of operation.

Fig. 5(*c*) is of particular interest: the insulator was taken from a valve constructed with one of the anode plates made of platinum and the other made of batch-B nickel, after 2 500 hours of operation under output conditions. The anode-to-grid capacitance had reached a fairly steady value of 0·060 pF. The black deposit is clearly visible, but only in the region in front of the nickel plate.

(5.2) Identification of Material found on Mica Insulators

The thinness of the black deposit made identification very difficult, and both spectrometric and electron-diffraction methods failed to yield any useful information. It was found that the deposit could be removed in flakes by gentle washing in water, but it was not attacked by the common acids or by aqua regia. The film was unaffected by firing the insulator at 500° C in dry hydrogen, but similar treatment in air produced complete removal in a short time. Finally, the deposits from ten insulators were removed and the aggregate was fired in wet nitrogen; the products of combustion were found to contain a significant quantity of carbon monoxide by the palladium-chloride test. These results indicate that the material of the deposit must be carbon.

(5.3) Discussion

It must be concluded that the deposit found on the mica insulators is the direct cause of the enhanced anode-to-grid capacitance; a black deposit always accompanies capacitance growth (provided that the valve has been operated long enough), and, from Fig. 4, conductivity of the material is detected where it reaches the glass bulb.

It can also be concluded that the material of the film is predominantly carbon, although it has not been proved to be carbon alone. Apart from the chemical evidence, this conclusion is consistent with the remarkable stability of the anode-to-grid capacitance when the valve is exposed to the atmosphere.

It is logical to link the above findings with the life-test results and postulate that carbon impurity in the anode is transferred to the insulators, but if this does indeed happen, some mechanism must be propounded. It is reasonable to assume that the carbon reaches the surface of the anode by diffusion at a rate dependent on anode temperature. Results obtained by Lander, Kern and Beach[6] show that the rate of diffusion of carbon in nickel at 350° C is about 500 times that at 250° C. Since the time for the onset of capacitance growth is about 100 hours for output valves with batch-D unfinned anodes, it follows that capacitance growth on input valves should not be detected before about 50 000 hours under the most favourable conditions, which agrees with experimental fact.

To explain how carbon leaves the surface of the anode is much more difficult. The position and sharpness of the boundaries of the deposit are consistent with the material having moved rectilinearly from the inner surface of the anode. There is no indication that the movement is ionic, because the varying potential of the mica would then have disturbed the shape of the pattern.

There is also no evidence that the temperature of the mica surface has influenced deposition. Evidently, any theory based on simple evaporation from the anode must fail, because the

deposit is confined to the mica on the inside of the anode, and all the evidence indicates that carbon is removed from the anode by a process resembling sputtering. There is, however, no known evidence for electron sputtering of any material at these voltages, except for a note by Jacobs[7] which does not seem to have been confirmed.* Despite this, it is still possible that, when it exists as an impurity in nickel, carbon has the property of being removed by electron bombardment. Alternatively, the carbon may appear on the surface of the anode in combined form, such as an adsorbed layer of carbon monoxide, which dissociates under electron bombardment, with the emission of the constituent elements. Evidence that electron bombardment can dissociate a compound and release its atoms as neutral particles has been obtained by Moore.[8]

(6) CONCLUSION

Allowing that there is still room for a clearer understanding of the chemical processes responsible for capacitance growth, it is nevertheless evident that the essential features of the phenomenon have been established. The anode plays the most important role and may well influence the general behaviour of a valve more than is generally supposed; from a mere inspection of an insulator bearing a black deposit it is clear that the cathode must also receive a share of the carbon from the anode.

It is believed that the phenomenon has been discovered on account of a fortunate combination of circumstances which has not previously arisen. In the first place, cathode evaporation will probably mask the phenomenon in valves with active cathode cores, but will not do so when platinum cores are used.[9] Secondly, in order to mitigate the effects of cathode evaporation, the mica insulators of active-core valves are normally coated with magnesium oxide, whereas the insulators in type 6P12 valves are always uncoated, because such a coating has been found to be incompatible with the highest levels of cathode emission. Finally, hydrogen-firing of valve components is a common commercial process.

* Attempts have also been made to detect simple sputtering by life testing valves having anodes of pure platinum coated with pyrolytic carbon, but no growth of anode-to-grid capacitance has been observed.

(7) ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in the paper.

(8) REFERENCES

(1) Child, M. R.: 'Notes on a Source of Intermittent Noise in Oxide-Cathode Receiving Valves', *Proceedings I.E.E.*, Paper No. 2140 R, September, 1956 (**103** B, p. 667).

(2) James, E. G., and Humphreys, B. L.: 'Resistive Films in Valves', *Wireless Engineer*, 1949, **26**, p. 93.

(3) Howe, G. W. O.: 'The Behaviour of High Resistances at High Frequencies', *ibid.*, 1935, **12**, p. 291.

(4) Howe, G. W. O.: 'The Behaviour of Resistors at High Frequencies', *ibid.*, 1940, **17**, p. 471.

(5) Reynolds, F. H., and Rogers, M. W.: 'A New Method for the Detection of Thin Conducting Films in Thermionic Valves', *Proceedings I.E.E.*, Paper No. 2335 R, May, 1957 (**104** B, p. 337).

(6) Lander, J. J., Kern, H. E., and Beach, A. L.: 'Solubility and Diffusion of Carbon in Nickel: Reaction Rates of Nickel-Carbon Alloys with Barium Oxide', *Journal of Applied Physics*, 1952, **23**, p. 1305.

(7) Jacobs, L.: 'Electron Bombardment as a Means of Material Transfer', *Nature*, 1946, **157**, p. 586.

(8) Moore, G. E.: Private communication.

(9) Metson, G. H.: 'A Study of the Long-Term Emission Behaviour of an Oxide-Cathode Valve', *Proceedings I.E.E.*, Paper No. 1790 R, April, 1955 (**102** B, p. 657).

621.315.615.9

Paper No. 2386 M
Sept. 1957

THE ELECTRIC STRENGTH OF SILICONE LIQUIDS

By T. J. LEWIS, M.Sc., Ph.D., Associate Member.

(*The paper was first received 26th October, in revised form 31st December, 1956, and 3rd April, 1957, and in final form 29th April, 1957.*)

SUMMARY

Controlled measurements of the d.c. electric strength of certain pure liquid poly-dimethyl siloxanes or silicones are reported. The liquids tested were the four lower members of the D.C. 200 series of silicone fluids having viscosities of 0·65, 1, 1·5 and 2 centistokes. The measurements were made under conditions identical with those employed for earlier experiments on pure *n*-paraffins which have a similar molecular structure, so that useful comparisons could be made. The use of a hydrogen-thyratron circuit to by-pass the discharge energy when breakdown occurred was found to be essential, since the discharge products, which probably included silica, were appreciably greater than for the *n*-paraffins.

Like that of the *n*-paraffins, the strength of the silicones, which is of the order of 900 kV/cm, was found to increase with chain length and with decreasing test-gap. The liquid of 0·65 centistokes viscosity was also tested over the temperature range −55° C to 85° C, and it was found that the strength increased with decreasing temperature, the increase being most marked at temperatures below about 20° C. This behaviour is similar to that of the *n*-paraffins already tested and cannot be ascribed to density changes alone.

(1) INTRODUCTION

The silicone fluids form a large class of liquid dielectrics having a very wide range of viscosities and being characterized by a low viscosity-temperature coefficient, a high degree of heat stability, oxidation resistance and general chemical inertness.[1] Thus they are attractive for use in electrical apparatus and, although expensive, have found several spheres of usefulness, e.g. in liquid-filled capacitors. They also have good dielectric properties, with a low power factor over a frequency range up to 100 Mc/s and a high resistivity (10^{14} ohm-cm). Concerning their electric strengths, the reported values for D.C. 200 type silicone liquids range from 250 to 500 volts/mil, which are comparable with test values for a normal transformer oil. The values given above are not likely to represent the ultimate strength that could be achieved with these liquids. Attempts to obtain this ultimate strength under controlled laboratory conditions are interesting because the results will indicate whether the values quoted above are reasonable. They will also provide valuable comparisons with results already obtained for *n*-paraffin liquids, which have an analogous chemical structure, providing interesting additional information concerning the relationship between electric strength and density, molecular structure, and viscosity.

The particular object of the present investigation was to measure the d.c. electric strength of certain low-viscosity D.C. 200 silicone liquids under conditions of high purity, similar to those used for earlier measurements[2] on pure hydrocarbon liquids, so that useful comparisons might be made. Unfortunately, only four liquids were available with sufficient initial chemical purity, but these provided sufficient data for reliable deductions to be made.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Lewis is in the Electrical Engineering Department, Queen Mary College, University of London.

(2) CHEMICAL STRUCTURE

The D.C. 200 silicone fluids in general consist of mixtures of linear poly-dimethyl siloxanes of the structure represented in Fig. 1(*a*), in which *n* is the number of dimethyl-siloxane units.

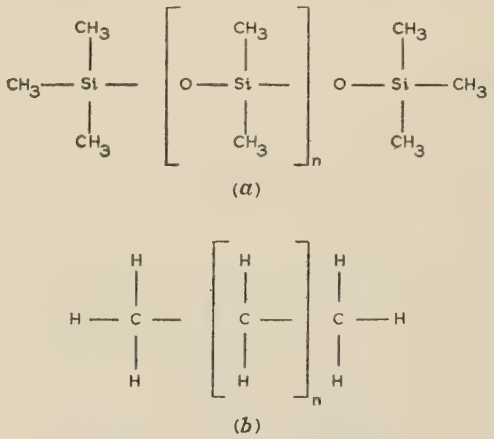


Fig. 1.—Molecular structures.
(*a*) Open chain polydimethyl siloxane (silicone).
(*b*) Normal paraffin.
Length of molecule determined by integer *n*.

In the higher polymers *n* can be as great as 2000, corresponding to viscosities up to 10^6 centistokes. The first four members of the series with $n = 0$, 1, 2 and 3 have viscosities of 0·65, 1, 1·5 and 2 centistokes, respectively, and it is these which can be obtained with sufficient purity. Liquids of higher viscosities are mixtures of various polymers and were not considered in the present experiments on that account. The relevant data concerning the first four members of the series are given in Table 1.

The analogous straight-chain *n*-paraffin liquids have structures as shown in Fig. 1(*b*), where *n* refers to the number of CH_2 groups in the molecule. These hydrocarbons are liquid at room temperatures only when *n* is between 3 and 15, the usual liquid on which extensive electric-strength measurements have been performed being *n*-hexane, for which $n = 4$. For a given value of *n*, the silicone molecule is much larger than the *n*-paraffin, the Si – O distance being 1·6Å and the C – C distance 1·54Å, whilst the side group dimensions are for Si – CH_3 greater than 2Å, and for C – H about 1Å (Fig. 1). The molecular weight of a silicone is also much greater than that of the corresponding hydrocarbon, so that for equal densities there will be fewer silicone than *n*-paraffin molecules per cubic centimetre. Data on the *n*-paraffin pentane are provided in Table 1 and should be compared with those for the pentamer, since both of these liquids have a chain length corresponding to $n = 3$ in Fig. 1.

(3) TEST PROCEDURE

The test procedure followed closely that reported earlier for measurements on pure hydrocarbon liquids.[2] The test cell formed part of a totally enclosed glass system in which the

Table 1

PHYSICAL PROPERTIES OF D.C. 200 SILICONE LIQUIDS

Name (polymer size)	Dimer	Trimer	Tetramer	Pentamer	*n*-paraffin pentane
Chain length (*n*, Fig. 1)	0	1	2	3	3
Boiling point, °C	99·5	152	192	230	36
Melting point, °C	−68	−86	−76	−84	−130
Viscosity, centistokes at 25° C	0·65	1·04	1·53	2·06	0·367
Specific gravity 25°/25° C	0·761	0·818	0·852	0·871	0·626
Molecular weight	162	236	310	384	72

The *n*-paraffin pentane is included for comparison with the pentamer which has the same chain length.

silicone liquids could be cleaned. The cleaning processes consisted of drying over sodium, reflux and slow distillation at reduced pressure to remove absorbed gas, and filtration through a sintered glass filter to remove dust particles having a size greater than about 10^{-3} mm. The liquid was then admitted to the test cell whence it could be returned and the cleaning cycle repeated. Before admitting the silicone liquids, the glass system was cleaned and degreased by circulating *n*-hexane of analytic purity round the system. This was then discarded and the whole system evacuated to remove the last traces of hexane.

The test cell and electrode system were identical with those used earlier, so that there is a reliable basis for comparing the results. The electrodes consisted of 1 cm diameter spheres of either chromium-plated brass or aluminium, and these were buffed until microscopic observation revealed no flaws or scratches on the test areas. They were then rinsed in dust-free hexane before insertion in the cell. Zero gap setting using a simple triode valve circuit and gap adjustment by means of an optical lever were the same as in the previous account,[2] so that the gap could be set to within $0{\cdot}25 \times 10^{-4}$ cm.

The high-voltage d.c. supply was carefully smoothed and capable of very fine control, and a stable 72-megohm resistor placed directly across the test gap was used for voltage measurement in conjunction with a valve voltmeter operating from a tapping on this resistor. It was considered that the overall voltage measurement was accurate to within $\pm 1{\cdot}5\%$, but reproducibility and long-term stability were much better than this.

Since a number of breakdown measurements were to be made in each sample of liquid, it was important that the discharge when breakdown occurred should be reduced to a minimum. It was noticed that the quantity of discharge products resulting from a breakdown was much greater for the silicones than for the hydrocarbons. After a series of discharges, opaque deposits could be observed in the high-field regions of the test gap, probably consisting of silica and perhaps carbon. Water, hydrogen and carbon dioxide are also likely to be given off when a discharge occurs.[1] Finely divided silica deposits would produce local discharges and a lowering of the breakdown strength of the gap. To reduce these deposits and to avoid excessive damage to the electrodes, an efficient diverter circuit involving both a series limiter hard valve and a parallel diverter thyratron was employed. A circuit of this type has been described by Saxe and Lewis.[3] With its aid, as many as 70 consecutive tests were possible before damage to the system necessitated fresh electrodes and liquid. Reliable estimates of the breakdown strength were therefore possible.

(4) MEASUREMENTS

Three separate tests were performed on samples of each of the silicone liquids to establish the effectiveness of the cleaning procedure adopted. In the first test, the commercial liquid was distilled and filtered once but without degassing before admission to the test cell. For the second, the liquid was recirculated, refluxed at reduced pressure to remove dissolved gas, and then distilled and filtered before passing to the test cell. In the final test, the liquid was dried over sodium for 24 hours and then degassed, distilled and filtered as for the second test. The higher-viscosity liquids, having higher boiling points and vapour pressures, required to be distilled at elevated temperatures, but the dimer (0·65 centistokes) had to be cooled during degassing in order to reduce loss by evaporation. All measurements in this series, however, were made at room temperature and atmospheric pressure, using chromium electrodes which were examined microscopically and thoroughly cleaned between tests. Each test consisted of a set of 10 breakdowns at a series of gap settings between 2×10^{-3} cm and 9×10^{-3} cm, and was preceded by a few conditioning discharges in which the strength rose to a consistent value. Usually only one or two of these discharges were required and never more than five. Similar conditioning effects had been found with the *n*-paraffins. In general, it was found that the strength for a particular gap setting could be obtained with sufficient accuracy from the mean of 10 measurements taken at half-minute intervals, the voltage being increased slowly to breakdown within that time. Comparison with *n*-hexane under similar conditions indicated, however, that the coefficient of variation in a long series of measurements (80 breakdowns with given sample and electrodes) was greater for the silicones, being of the order of 5%, in comparison with 1%. This was attributed to the heavier deposits which occurred on breakdown with the silicones. By limiting the number of measurements to 10 at each gap setting, reproducible results could be obtained and reliable deductions made from these. The reliability of the tests was also checked by repeating the earlier measurements with *n*-hexane.[2] The new estimates of the electric strength were in close agreement with the old.

Table 2 gives typical results for the silicones at a gap of 8×10^{-3} cm, and shows that there is a significant increase in

Table 2

EFFECT OF CLEANING ON BREAKDOWN STRESS (MV/CM) FOR A 8×10^{-3} CM GAP

Test procedure	D.C. strength MV/cm			
	Dimer	Trimer	Tetramer	Pentamer
(i) Distillation and filtration	0·93	0·98	1·04	1·10
(ii) Distillation, filtration and degassing	0·96	1·00	1·11	1·16
(iii) Drying, distillation, filtration and degassing	0·98	1·03	1·14	1·17

strength as a result of the cleaning and drying processes. Values of electric strength obtained from the third test were taken as those of the clean degassed liquid at room temperature, and yielded linear plots for the variation of breakdown voltage with gap setting, as shown in Fig. 2. The characteristics of the

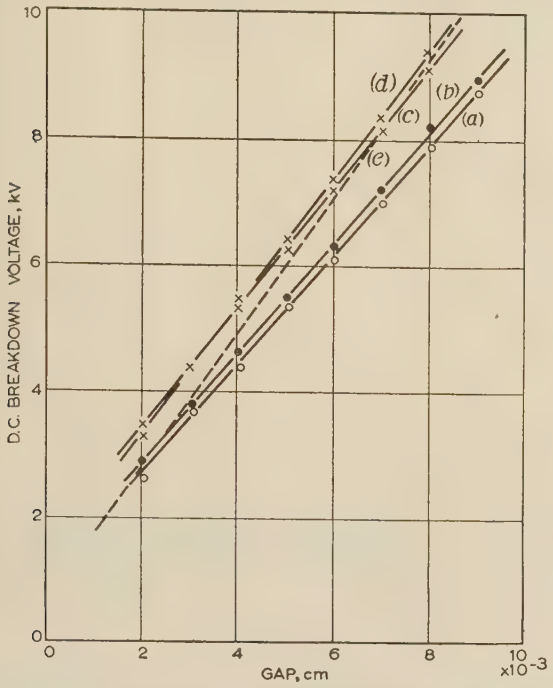


Fig. 2.—Dependence of breakdown voltage on gap setting.
(*a*) Dimer. (*b*) Trimer. (*c*) Tetramer. (*d*) Pentamer. (*e*) *n*-hexane.

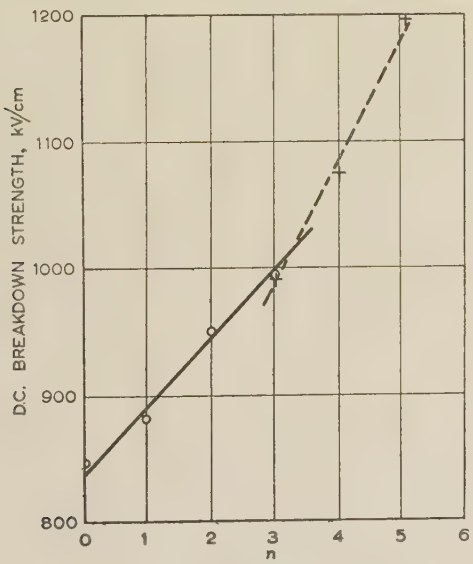


Fig. 3.—Variation of d.c. electric strength with molecular chain length (*n*).
The strength is calculated for a large gap.
——— Silicones. – – – *n*-paraffins from Reference 2.

silicones and *n*-paraffins are similar, as can be seen by comparing the results for the silicones with that obtained for *n*-hexane which is included in Fig. 2, and also with earlier results for the hydrocarbons.[2] The strengths of all the liquids increase with decreasing test gap. The slopes of the characteristics can be taken as the strengths at large gap setting, and therefore represent the d.c. electric strength of the bulk of the liquid. These values are utilized in Figs. 3 and 4, which show that the strength at room temperature increases with chain-length density and viscosity.

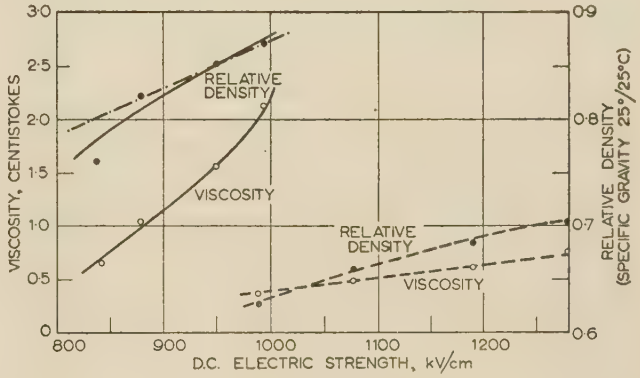


Fig. 4.—Variation of d.c. electric strength with viscosity and density.
——— Silicones. — — — *n*-paraffins, pentane to octane, from Reference 2.
— · — Extrapolated density/electric-strength curve of the three higher silicone polymers.
The strengths are calculated for a large gap.

If the electrodes were changed from chromium to aluminium the above results were not altered, which again is in agreement with the measurements on hydrocarbons.

(4.1) Effect of Temperature

The effect of temperature on the strength should be interesting not only because of the inherent heat stability of these liquids, but also because the dependence of viscosity on temperature is much less than it is for the *n*-paraffins. It is therefore possible that the properties that result in a small variation in viscosity may also produce a small change in electric strength.[4,5] Unfortunately, circumstances limited measurements to the dimer hexamethyldisiloxane, and the results are shown in Fig. 5. The

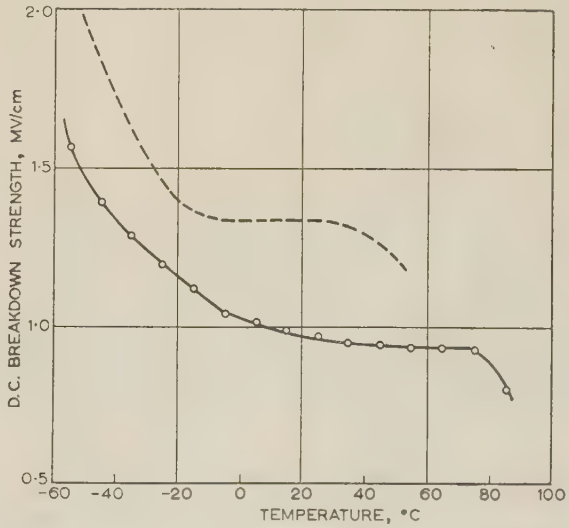


Fig. 5.—Dependence of d.c. electric strength on temperature.
——— Silicone (dimer) for a gap of 5×10^{-3} cm.
— — — *n*-paraffin (hexane) for a gap of 3×10^{-3} cm.

gap was set at $5\frac{1}{2} \times 10^{-3}$ cm and maintained constant at this value by applying a predetermined correction to allow for

expansion of the electrode system.[2] A thermocouple within the lower-electrode support permitted temperature measurement to within 1° C, the test cell being immersed in a liquid bath in a vacuum vessel to stabilize the temperature. A complete temperature characteristic required a large number of tests to be made on one sample of liquid, so in order that the effects of discharge damage should not affect the results seriously, each representative point in Fig. 5 was determined from 5 and not 10 determinations. None of these points lies far from the smooth curve shown, which indicates a reasonable accuracy.

The strength remains fairly constant over a considerable portion of the upper temperature range, and the rapid drop at about 85° C may indicate the beginning of ebullition in the liquid, since the boiling point is 99·5° C. Fig. 5 shows that the rate of increase of strength rises as the temperature decreases in the low temperature range, and this behaviour is similar to that reported earlier for the *n*-paraffins[2], of which the typical result for *n*-hexane at a gap of 3×10^{-3} cm is included in the diagram. Further discussion taking into account the effect of density changes will be found in Section 5.2.

(5) DISCUSSION OF RESULTS

(5.1) Dependence on Liquid Properties

The general behaviour of the low-viscosity silicone liquids follows closely that of comparable *n*-paraffins, except that the discharge products have more influence in the former case. It is possible that silica deposits, which are, of course, absent in the hydrocarbons, cause the strength to fall when many breakdowns occur. Both types of liquid have breakdown-voltage/gap-setting characteristics which are linear and do not pass through the origin when extrapolated (Fig. 2). These results suggest that it may be useful to consider that the breakdown mechanism in these silicone fluids is of the same type as in the hydrocarbons,[6, 7] and that we should discuss the breakdown strength in terms of molecular properties such as density, chain length, etc., making comparisons with the *n*-paraffins. Electrode effects, which are not marked in any case, are rendered unimportant by using values corresponding to a large gap determined from Fig. 2.

The electric strength increases with chain length (i.e. with *n*) as shown in Fig. 3, and this resembles the behaviour of the *n*-paraffins, but it is not likely that chain length alone can determine the electric strength. The silicones have significantly lower strengths than the *n*-paraffins of corresponding density or viscosity (Fig. 4). From Fig. 4 it would appear that neither viscosity nor density alone is the fundamental parameter determining the strength. An investigation of those basic molecular properties which could determine the strength and account for the differences between the silicones and the *n*-paraffins would involve concepts of the breakdown processes[6, 7] which cannot be discussed here. It should be noted that the dimer has a considerably higher strength than would be expected from the extrapolated density/electric-strength curve for the other polymers shown in Fig. 4. This exceptional behaviour of the dimer has been noted in measurements of other properties also.

(5.2) Electric Strength and Temperature

If the strength is related to density, it is obvious that density changes due to change in temperature will cause a corresponding change in strength. Since the density increases as the temperature is lowered, the increase in strength at low temperature (Fig. 5) could be accounted for. By plotting strength/density against temperature, any variation due to density can be annulled. The result is given in Fig. 6, which shows a region of little change

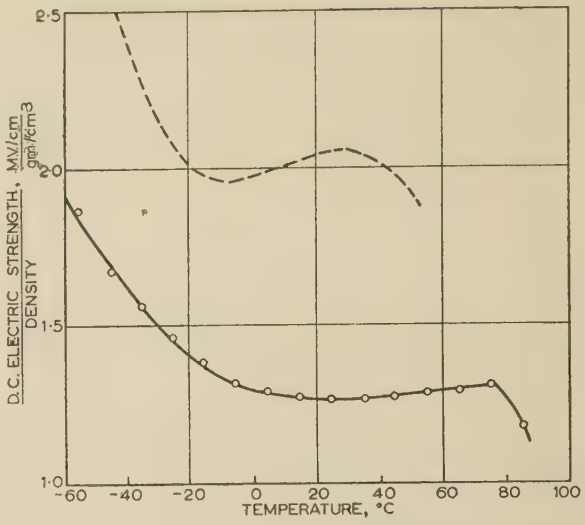


Fig. 6.—Variation of the ratio electric-strength/density with temperature, calculated from the results of Fig. 5.

——— Silicone (dimer). — — — *n*-paraffin (hexane).

from 10° C to 75° C but still a range in which the strength increases at lower temperatures. A somewhat similar result has already been found for the *n*-paraffins,[8] and a typical curve for *n*-hexane is also shown in Fig. 6 for comparison.

(6) CONCLUSIONS

Although unfortunately the test programme has been a limited one, there is sufficient evidence to show that the low-viscosity poly-dimethyl siloxanes have breakdown strengths under comparable conditions which are lower than those of the corresponding *n*-paraffins. This is true when the comparison is made on the basis of either density or viscosity (Fig. 4). As far as can be ascertained from the tests, the breakdown characteristics appear to be of the same form in both liquids, but the discharge products are more deleterious in the silicones. This latter is probably associated with the formation of silica.

The strengths of the bulk liquid for the four types tested ranged from 840 kV/cm to 990 kV/cm. The value of 840 kV/cm for hexamethyldisiloxane is in close agreement with the isolated result of 820 kV/cm reported by Musset, Nikuradse and Ulbrich.[9] As far as is known there are no other reported values for the electric strength of these silicones under controlled conditions.

(7) ACKNOWLEDGMENTS

The author wishes to thank Mr. D. Brook, who ably carried out the experiments reported here, and also acknowledges gratefully the award of a Turner and Newall Fellowship by the University of London which enabled this work to be performed.

(8) REFERENCES

(1) Bass, S. L., Hunter, M. J., and Kauppi, T. A.: 'Silicones as Electrical Insulating Materials', *Transactions of the Electrochemical Society*, 1946, **90**, p. 255.

(2) Lewis, T. J.: 'Electrical Breakdown in Organic Liquids', *Proceedings I.E.E.*, Paper No. 1488 M, March, 1953 (**100**, Part IIA, p. 141).

(3) Saxe, R. F., and Lewis, T. J.: 'Measurement of Statistical Time Lag of Breakdown in Gases and Liquids', *British Journal of Applied Physics*, 1955, **6**, p. 211.

(4) SALOMON, T.: Discussion on 'Electric Strength and Breakdown Mechanisms', *Proceedings I.E.E.*, 1953, **100**, Part IIA, p. 183.
(5) LEWIS, T. J.: *ibid.*, p. 186.
(6) LEWIS, T. J.: 'Mechanism of Electrical Breakdown in Saturated Hydrocarbon Liquids', *Journal of Applied Physics*, 1956, **27**, p. 645, and references cited therein.
(7) SHARBAUGH, A. H., CROWE, R. W., and COX, E. B.: 'Influence of Molecular Structure upon the Electric Strengths of Liquid Hydrocarbons', *ibid.*, 1956, **27**, p. 806.
(8) LEWIS, T. J.: 'Electric Strength and Liquid Structure', *ibid.*, 1957, **28**, p. 503.
(9) MUSSET, E., NIKURADSE, A., and ULBRICH, R.: 'Über die Durchbruchsfeldstärke in dielektrischen Flussigkeiten verschiedener molekularer Zusammensetzung', *Zeitschrift für Angewandte Physik*, 1956, **8**, p. 8.

DISCUSSION ON

'THE INSTRUMENTATION OF A 14IN EXPERIMENTAL ROLLING MILL'*

NORTH-EASTERN RADIO AND MEASUREMENT GROUP, AT NEWCASTLE UPON TYNE, 30TH OCTOBER, 1956

Mr. G. Ovens: The instrumentation systems described in the paper will be useful in assisting to improve the operation of mills on continuous production. The authors' claim that they serve as useful models for industrial instruments is already being borne out in practice.

The gaugemeter shows the greatest promise of being an effective tool for production, both on reversing and tandem mills, since it gives a measurement of gauge at the point where corrective action can be taken. Any gauge which measures the thickness a few feet after the point at which it can be regulated subjects the control system, whether manual or automatic, to a transport lag which cannot readily be overcome. This is certainly so in a continuously-operating regulating system.

Sections 5 and 6 of the paper, covering automatic control, are of particular interest for all types of mill. A recent tandem cold reduction mill has been equipped with Ward Leonard screws, so that an automatic gauge-control system operating by continuous action can be applied to the entry stand. On a tandem mill a control system operating from a gaugemeter, such as the authors describe, could do much to maintain gauge and to correct for gauge discrepancies in the hot hand by acting on the screws of the entry stand. 'Continuous' would be preferred to 'on-off' action on a production mill in almost continual operation because maintenance of contactors with an 'on-off' scheme would be very high. For the control of finished gauge from a tandem mill, a gaugemeter at the last stand could be connected to control the stand motor and thus to control the interstand tension and thereby the gauge.

In Section 4, covering the performance of the measuring system, no mention is made of speed-dependent errors which might upset the linear relationship $h = S + F/M$. This is important on a production mill, for, while it is relatively easy to maintain gauge while running steadily at high speed, it is exceedingly difficult to do so while accelerating or decelerating. A gaugemeter which would give, for manual or automatic control, an exact reading of departure from gauge during acceleration and deceleration, would be of use. For this it would be necessary to know the speed-dependent characteristic of the combination of gaugemeter and stand and to feed in a compensation.

The tension meter, whether used merely for indication or for automatic control, will not properly act as a tension-measuring device under conditions of speed change. While this is probably not important for many of the studies which have to be conducted on an experimental mill, in industrial applications, compensation would need to be injected into the system to give a reasonable 'dynamic' accuracy, which, in fact, needs to be done in schemes using current as the intermediate regulated quantity.

Messrs. S. S. Carlisle and **G. W. Alderton** (*in reply*): Mr. Ovens's remarks are interesting and support many of the claims made in the paper.

The speed-dependent errors of the gaugemeter were not studied in detail, but it is not considered that such errors will be significant, because most of the gauge change of speed appears to be due to a change of friction in the roll gap, thus giving a different percentage reduction for otherwise similar rolling conditions. The gaugemeter would only show a speed-effect error if there were to be a significant change in lubricating film thickness or of the position of the roll necks in their bearings, as might be caused by changes in the conditions of hydrodynamic lubrication with speed change. Although such influences can be appreciable in some mills they are not now thought to be very significant on a high-speed cold rolling mill with good roll bearings.

We agree that the tension measuring device described is subject to error when coiler speed is changing. The system was only used as an interim measure. The three-roller system* for tension measurement is generally recommended as being more satisfactory in all respects.

* CARLISLE, S. S., and ALDERTON, G. W.: Paper No. 1933 M, November, 1955 (see 103 B, p. 360).

* SIMS, R. B.: *Engineering*, 1952, **174**, p. 262.

621.387:621.3.032.12:546.293

Paper No. 2394 R
Sept. 1957

INFLUENCE OF ARGON CONTENT ON THE CHARACTERISTICS OF GLOW-DISCHARGE TUBES

By F. A. BENSON, D.Eng., Ph.D., Associate Member, and E. F. F. GILLESPIE, M.Eng., Graduate.

(*The paper was first received* 19*th February, and in revised form* 23*rd May*, 1957.)

SUMMARY

Measurements have been made to determine the effects of varying the argon content of glow-discharge stabilizer tubes on the striking and running voltages, the magnitudes and durations of the initial drifts, the running-voltage/temperature and running-voltage/current curves and the impedance/frequency and noise characteristics. Neon-filled and helium-filled tubes, with argon contents varying from zero to 3·5% and having cerium cathodes, have been specially manufactured for the investigations. The results of the work are presented here and discussed.

LIST OF PRINCIPAL SYMBOLS

x = Position in cathode region of glow discharge measured from the edge of the cathode-fall space remote from the cathode.
V = Potential at point x.
E = Electric field strength.
E_c = Electric field strength at cathode.
E_1 = Electric field strength at the transition point between high- and low-field regions in the cathode-fall space.
ρ = Total charge density.
ρ_+ = Positive-ion charge density.
ρ_- = Electron charge density.
J = Total current density.
J_+ = Positive-ion current density.
J_- = Electron current density.
v_+ = Positive-ion velocity.
v_- = Electron velocity.
v_c = Positive-ion velocity at cathode.
v_d = Positive-ion velocity at point $x = d$.
$\gamma = J_-/J_+$.
p = Gas pressure.
k_1 = A coefficient relating v_+ and E/p in the low-field region $x = 0$ to $x = d$.
k_2 = A coefficient relating v_+ and $(E/p)^{\frac{1}{2}}$ in the high-field region $x = d$ to $x = d_c$.
d_c = Length of cathode-fall space.
d = Extent of low-field region in cathode-fall space.
τ_1 = Transit time of ions in region $x = d$ to $x = d_c$.
τ_2 = Transit time of ions in region $x = 0$ to $x = d$.
τ = Transit time of ions across cathode-fall space $= \tau_1 + \tau_2$.
V_n = Normal cathode fall.

(1) INTRODUCTION

Glow-discharge tubes are widely used as simple voltage regulators and as reference elements in thermionic-valve stabilizers, but the characteristics desirable in such tubes vary with their usage: a stabilizer tube should have a large current range, a low impedance and a good voltage/current characteristic, while in a reference tube the variations in running voltage, at constant current, with time and temperature are of utmost importance. In recent years a great deal of work[1–5] has been carried out to determine the performances of many types of tube; further work, covering a much wider field,[6] has included studies of variations in striking and running voltages, step and hysteresis effects and the effects of temperature changes. Initial drifts of the tubes were also recorded, and some figures for voltage drifts to be expected during the life of a tube were included. In addition, some information has been published on the variations of extinction voltages[7] of tubes and on the characteristics observed with intermittent operation,[8] and further investigations on initial drifts and effects of ambient-temperature changes have been carried out more recently.[9–11] The a.c. impedance of a tube at high frequencies is considerably greater than that quoted in data sheets for low frequencies.[12–21] Measurements have also been made to determine the noise characteristics of tubes.[22–25] Most of the work referred to above was carried out with commercial stabilizer tubes, but the latest investigations[11,21,25] have been made with special tubes differing in cathode material, gas filling, gas pressure and construction. The argon content of these special manufactured tubes, however, varied from 0·3 to 2·5%, and the results obtained suggested that these variations played an important part in determining the tube characteristics, especially the impedance/frequency curves.

As a consequence of these observations it was decided to construct special tubes in which only the argon content was varied. It was realized that the manufacturing procedure was an important factor in determining the tube characteristics, and so the samples were made under exactly similar conditions and from identical materials. Sixteen batches, each containing six identical valves, were produced, eight batches being filled with helium and eight with neon; in each of these two groups the argon content was varied from zero to 3·5% in 0·5% steps. The gas pressure was 40 mm Hg in the helium-filled tubes and 50 mm Hg in the neon-filled ones. Each tube had a rod anode 1 mm in diameter surrounded by a cylindrical cathode 20 mm in diameter and 30 mm long; the cathode cylinder had mica sheets at each end to enclose the working portion of the tube, and the electrode structure was mounted in a glass envelope having a volume of about 45 cm^3. Cerium cathodes were used in all tubes, and the final ageing process was carried out carefully to ensure identical processing. Measurements were made on these tubes to determine the influence of argon content on the striking and running voltages, the magnitudes and durations of the initial drifts, the running-voltage/temperature and running-voltage/current curves, and the impedance/frequency and noise characteristics. The results of this work are presented here and discussed.

(2) MEASUREMENTS

The methods used for measuring the initial drift, running-voltage/temperature and running-voltage/current changes have already been described.[5,10] Each tube was fed in turn from a stabilized d.c. supply (the output voltage of which was constant to within 0·1%) through a suitable series resistor. A meter in the anode lead was used to measure the current, while two decade resistance boxes across the tube stepped down the

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Benson and Mr. Gillespie are in the Department of Electrical Engineering, University of Sheffield.

running voltage so that it could be measured, to within 10 mV, on a vernier potentiometer. The ambient temperature was varied by immersing the tube in an electrically-heated oil-bath.

The running voltage of each tube was recorded as the current was varied in steps, first from 5 to 40 mA and then from 40 to 5 mA. At each particular current the running voltage drifted somewhat, so time was allowed for stable values to be reached. It was noticed that the drifting period persisted for a considerable time at high tube currents and increased with increasing argon content. With tubes having argon contents of 3 and 3·5% the voltage drifted for 15–20 min at each current step. It was therefore decided to measure these characteristics for only two tubes from each group having the same argon content. Throughout the measurements of running-voltage/current characteristics the tubes were operating in air at an ambient temperature of about 20° C and the glass envelopes were shielded from draughts.

To obtain the running-voltage/temperature curves each tube was operated at 20 mA and the running voltage was recorded at temperature intervals of about 10° C in the range 20–90° C. Before each reading was taken the heater current was interrupted and the temperature was allowed to reach a steady value to achieve a constant running voltage. It has been pointed out[10] that the voltage drift with temperature of a tube in an oil-bath is somewhat different from that realized in practice, since with normal ambient-temperature variations the glass envelope of the tube would generally be hotter than its surroundings.

Initial drifts have been measured only for the various helium-filled tubes at a current of 20 mA, and for most tubes the drifts lasted about 10 min.

For impedance measurements it was found that the circuit previously used[21] became unreliable at frequencies approaching 100 kc/s, probably because of the effects of stray capacitances. In this circuit the tube under test was fed through a suitable resistor from a stabilized d.c. supply, a meter in the anode lead being used to set and maintain the tube current at the desired value. In addition, the tube was fed from an a.f. oscillator, one side of which was connected to the anode through a very large blocking capacitor and the other side to the cathode through a non-inductive resistor, across which a valve voltmeter was used to keep the superimposed alternating current at 1 mA. The alternating voltage drop across the tube was also measured with a valve voltmeter. The voltage-dropping resistance was high, but to ensure that its shunting effect across the tube could be neglected, a choke was placed in series with the d.c. supply. The alternating voltage and current were fed to the *Y*- and *X*-plates respectively of an oscillograph; the phase-angle difference between the voltage and the current was measured directly from the dynamic characteristic traced on the screen, and the tube impedance was thus divided into its real and imaginary components.

To overcome the unreliability of the original circuit at high frequencies the arrangement shown in Fig. 1 was employed during the present investigations. The new circuit has a common point at A, so that the phase-angle measured includes the effect of the resistance between A and B. To enable a direct value of tube impedance to be measured, the 10-kilohm a.c. load was placed on the positive side of the tube and a 50-ohm resistor was placed between A and B to give the current vector needed for the phase-angle determination. This 50-ohm load, which was obtained from a decade box, was made zero after the alternating current had been set at 1 mA r.m.s., so that the valve-voltmeter reading gave a direct indication of the impedance. The change in the alternating current due to this variation of resistance was small, since the impedance of the circuit was greater than 10 kilohms. To overcome any phase shift due to a difference

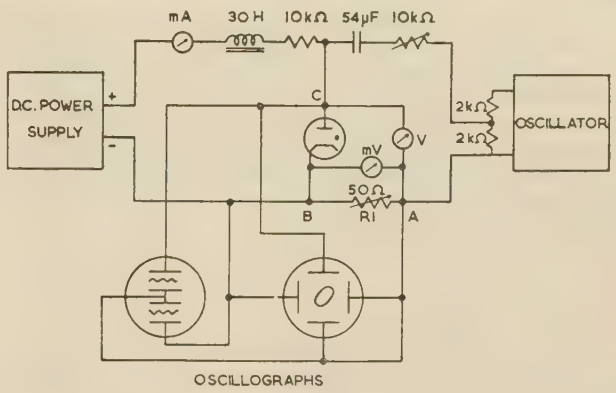


Fig. 1.—Circuit for measuring the a.c. impedance of glow-discharge tubes.

in the amplifiers, two identical amplifiers were used to form the ellipse, the second amplifier being obtained from an additional oscillograph. Since double-beam oscillographs were used, the second model was used to display the current and voltage waveforms.

It was found impossible to measure the impedance of tubes containing pure helium or pure neon, since persistent oscillations of the plasma were observed. The impedances of all helium-argon tubes were obtained over the range 20 c/s–100 kc/s at a tube direct current of 20 mA and a superimposed alternating current of 1 mA r.m.s. With neon-argon tubes internal oscillations were observed, and to minimize these the tube direct current was reduced to 5 mA and the alternating current to 0·5 mA r.m.s. The load resistance was then increased to 100 ohms to give an ellipse of convenient size. Even under these conditions it was found impossible to obtain values of phase angle and impedance above 1 kc/s. For neon tubes containing more than 2·5% argon no results could be recorded, owing to these persistent oscillations. At this stage some standard commercial tubes containing neon-argon mixtures with only 0·3% argon were examined and were found to behave normally, thus showing that argon content was an important parameter in producing internal oscillations.

The noise measurements were made with a calibrated amplifier-detector unit having a nearly perfect linear response over the range 10 c/s–100 kc/s. The circuit arrangement was almost identical with one previously employed,[25] but had to be adjusted to cope with the increased current range of the present tubes and the large differences between the striking and running voltages of certain batches of tubes. As with the impedance measurements, internal oscillations seriously limited the number of noise readings obtainable.

(3) RESULTS

The mean curves of the variations in striking and running voltages with changes in argon content are shown in Fig. 2; all running voltages were recorded for a tube current of 20 mA and all tubes were suitably aged after manufacture before being tested.

Figs. 3 and 4 show typical running-voltage/current curves for helium- and neon-filled tubes with varying argon contents.

Fig. 5 illustrates typical runnning-voltage/temperature characteristics of tubes filled with pure helium and a pure neon, while Figs. 6 and 7 show how this characteristic varies with argon content.

The initial drifts of two helium-argon tubes per batch are shown as a function of time in Fig. 8.

In Fig. 9 the impedance of helium-filled tubes has been plotted as a function of the argon content; curves have been plotted for

0·1, 1, 10, 20, 40, 80 and 100 kc/s, but only the mean values for each batch are shown. Fig. 10 shows the tube impedance, resistance and reactance and the phase angle as functions of frequency for tubes having a gas filling of 99·5% helium and 0·5% argon. Similar curves have been obtained for other argon contents but are not shown.

The results obtained from the noise measurements have not been plotted, since only small random changes were observed as the argon content was varied.

(4) DISCUSSION OF RESULTS

(4.1) Striking and Running Voltages

From Fig. 2 it is evident that the amount of argon included in the gas-filling of both helium and neon tubes has a great effect

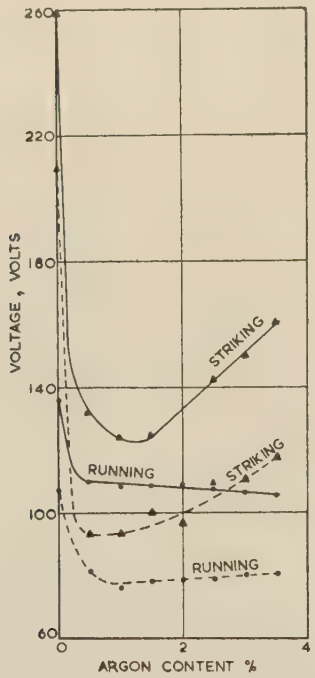


Fig. 2.—Striking and running voltages as a function of argon content.
—— Helium-filled tubes.
---- Neon-filled tubes.

on the striking voltage and an appreciable effect on the running voltage, reducing the former by about 100 volts and the latter by 20–30 volts. A minimum difference was obtained between these two voltages with an argon content lying between 1·0 and 1·5% for helium-filled tubes, and less than 0·5% for neon-filled tubes. It is interesting to note that, after the initial drop in running voltage due to the first addition of argon to the gas-filling, the further addition of argon continues to reduce the running voltage of helium tubes but increases it in neon-filled tubes once the minimum value obtained with about 1% argon has been reached. Jurriaanse, Penning and Moubis[26] have carried out investigations on molybdenum/neon-argon tubes, and they too found that a minimum running voltage was obtained with 1% of argon, while Meissner[27] also confirmed this for tubes with nickel cathodes. It should be noted that the three different types of tube referred to had different pressures: the cerium tubes used in the present investigations were filled to 50 mm Hg, the molybdenum tubes to 40 mm Hg, and the nickel tubes to only 7 mm Hg.

(4.2) Running-Voltage/Current Characteristics

It was found that with both helium- and neon-filled tubes containing no argon the running voltage fell as the tube current increased [Figs. 3(*a*) and 4(*a*)], the change in voltage over the current range being large (about 4%) for both types of tube. The voltage/current characteristics of the helium-filled tubes also exhibit considerable hysteresis, the characteristics for increasing and decreasing currents varying by more than 5 volts at some currents, as illustrated in Fig. 3(*a*).

The addition of argon to both helium- and neon-filled tubes not only lowered the running voltage, as already discussed, but also decreased the tube regulation, although steps in the characteristics gave certain tubes rather poor regulation figures. In general, tubes having helium-argon fillings exhibited larger hysteresis effects than those with neon-argon fillings. With the exception of tubes containing no argon, the voltage jumps in the running-voltage/current curves increased with increasing argon content. These steps were considerably smaller for neon tubes than for helium ones for the same argon content; they were generally sudden falls in the running voltage and occurred in the current range 5–20 mA for increasing current only.

Emeléus[28] has noted that, with mixtures of two inert gases, the one with the higher atomic weight and smaller ionization potential has a tendency to move towards the cathode end of the tube. Thus, argon, having a lower ionization potential than neon and helium, would have tended to drift towards the cathode and altered the conditions in this region according to the amount present. Since steps have been attributed, amongst other things, to the fact that the cathode surface in a tube is not perfectly homogeneous, it is perhaps not surprising that they should be influenced by argon content, but it is difficult to explain why helium- and neon-filled tubes behave so differently. The results might also be expected to depend on the cathode material.

(4.3) Running-Voltage/Temperature Characteristics

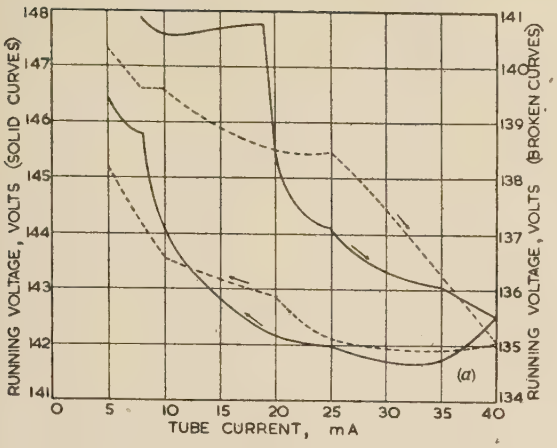
Tubes which contained no argon exhibited a negative temperature coefficient of running voltage very much greater than those which did, the change in voltage over the range 20–90° C being 1·1 volts for the pure-neon-filled tube and 1·5 volts for the pure-helium-filled tube (Fig. 5).

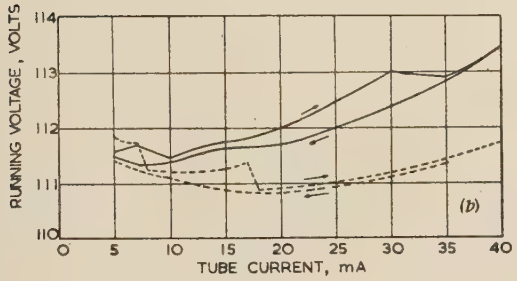
The effects of adding different amounts of argon to the tubes are indicated by Figs. 6 and 7; it will be seen that the voltage changes with temperature were very small, and that variations from tube to tube in certain batches were greater than variations from batch to batch. The argon content is, in fact, not very important in this respect, but it does seem that helium-filled tubes with an argon content of 1% have the minimum change in running voltage over the temperature range considered.

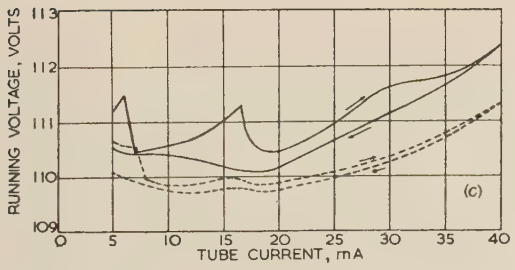
The steps in the curves for the helium-filled tubes containing 2·5% argon were probably due either to small changes of tube current or to changes of current at which steps appeared when the temperature was raised, because it was known that these tubes exhibited small steps in their voltage/current curves in the region of 20 mA.

For all neon-filled tubes the temperature coefficient was found to be negative in the range 20–40° C, and usually up to 80° C, and there was a tendency for it to increase slightly with increasing argon content. Neon-filled tubes with a low argon content (up to 1·5%) generally had temperature coefficients of almost zero over the range 40–70° C, becoming positive over the range 70–90° C.

Jurriaanse[29] explained theoretically the variation of running voltage with temperature and concluded that a negative temperature coefficient should be obtained. This coefficient was ascribed to small gas-density variations in the region of the







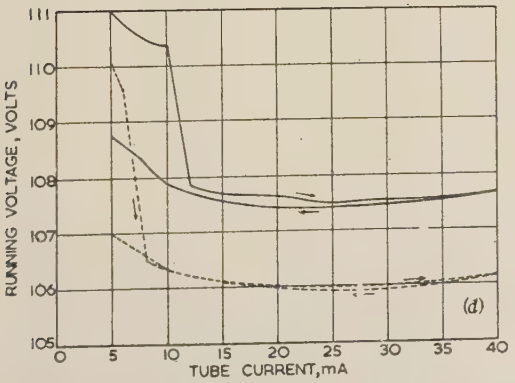
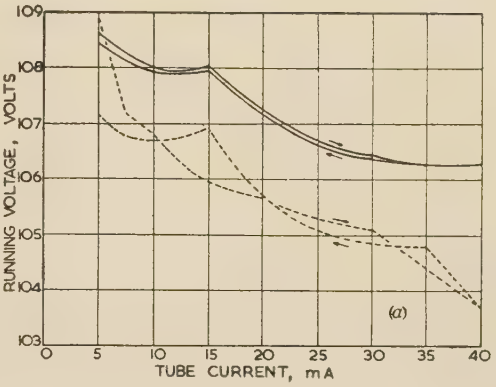
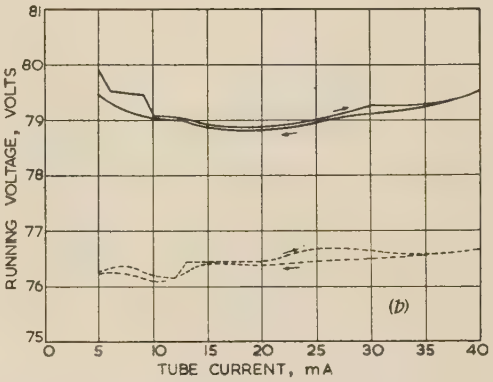
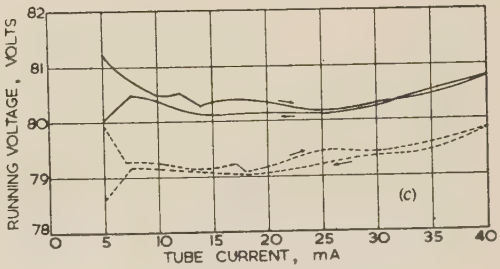


Fig. 3.—Running-voltage/current characteristics for helium-filled tubes.

(*a*) No argon.
(*b*) 0·5% argon.
(*c*) 1·5% argon.
(*d*) 3·5% argon.







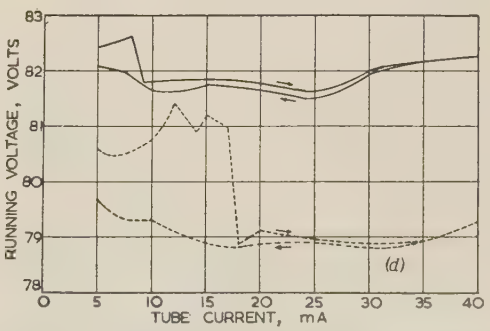


Fig. 4.—Running-voltage/current characteristics of neon-filled tubes.

(*a*) No argon.
(*b*) 1·5% argon.
(*c*) 2·5% argon.
(*d*) 3·5% argon.

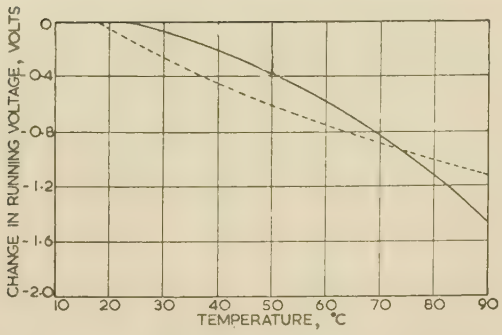


Fig. 5.—Voltage/temperature characteristics of tubes without argon.
——— Helium-filled tube.
- - - - Neon-filled tube.

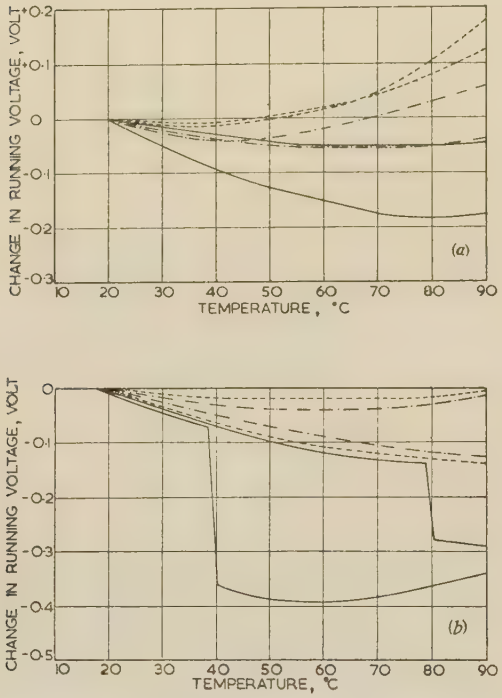
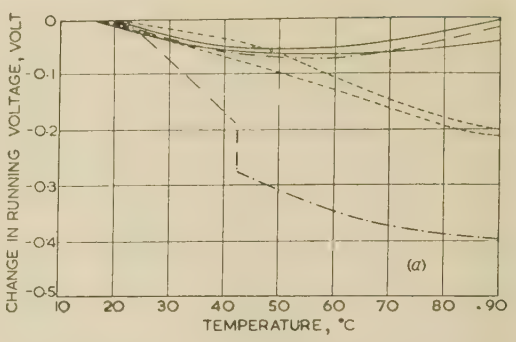


Fig. 6.—Voltage/temperature characteristics of helium-filled tubes.
(*a*) Argon contents up to 1·5%.
——— 0·5%.
—·— 1·0%.
- - - - 1·5%.
(*b*) Argon contents up to 3·5%.
——— 2·5%.
—·— 3·0%.
- - - - 3·5%.

cathode. It has been assumed[9, 10] that observed positive coefficients are caused by impurities and/or adsorbed gases.

(4.4) Initial Drifts

The initial-drift curves for helium-filled tubes shown in Fig. 8 indicate the effect of argon content in the gas filling. With tubes containing 0·5% argon, positive drifts were observed; increasing the argon content to 1·0% gave both positive and negative drifts, while further increase in argon content increased the negative drift. It is therefore concluded that the amount of drift depends on the argon content, but in all cases the drifting was almost complete about 3 min after switching on.



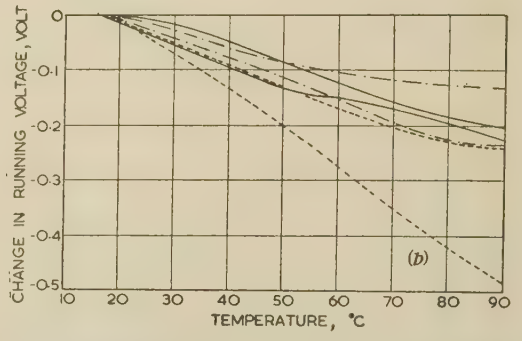


Fig. 7.—Voltage/temperature characteristics of neon-filled tubes.
(*a*) Argon contents up to 1·5%.
——— 0·5%.
—·— 1·0%.
- - - - 1·5%.
(*b*) Argon contents up to 3·5%.
——— 2·5%.
—·— 3·0%.
- - - - 3·5%.

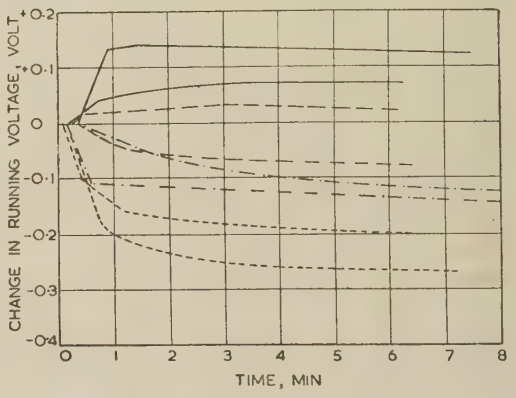


Fig. 8.—Initial drift of helium-argon tubes.
——— 0·5% argon.
— — 1·0% argon.
—·— 1·5% argon.
- - - - 2·5% argon.

(4.5) Impedance/Frequency Curves

Fig. 9 demonstrates clearly the effect of argon content on the impedance of a helium-filled tube: at 100 c/s the impedance fell from 65 to 35 ohms as the argon content increased from 0·5 to 3·5%, the corresponding figures at 10 and 100 kc/s being 450 to 360 ohms and 2100 to 760 ohms respectively; above 10 kc/s large changes of impedance occurred in the range 0·5–1·0% argon. From Fig. 9 it is concluded that, when tubes are required to operate in circuits where superimposed frequencies greater than 10 kc/s are present, the gas filling should contain as much

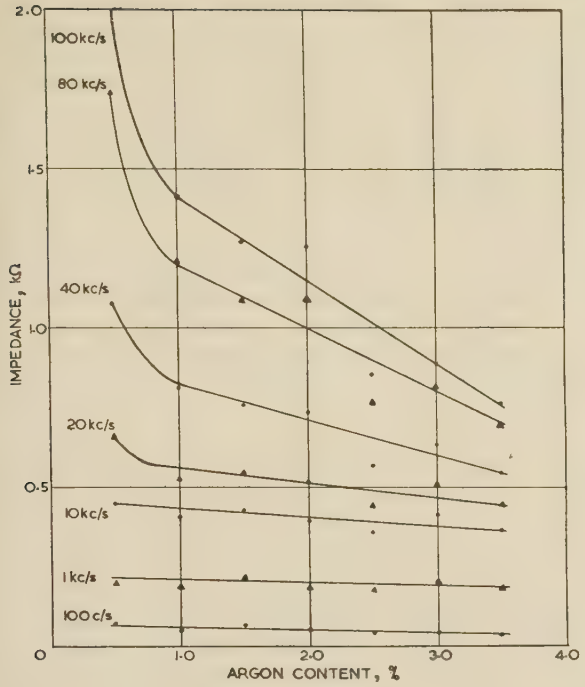


Fig. 9.—Variation of impedance with frequency and argon content of helium-filled tubes.

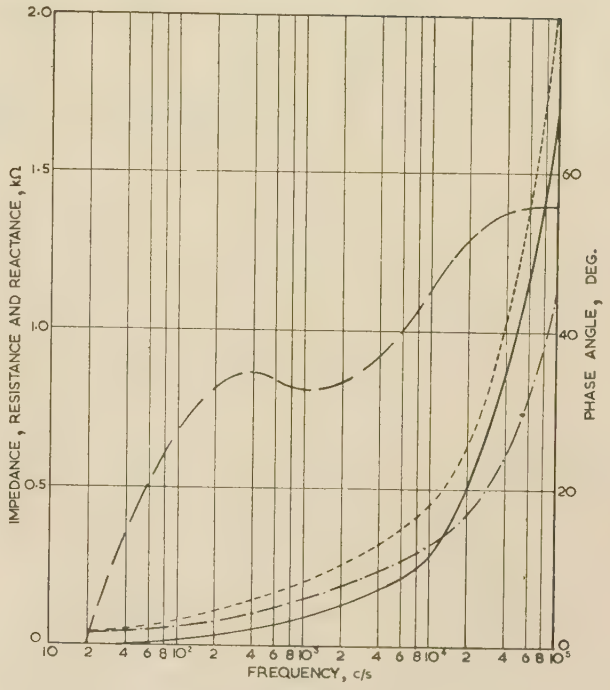


Fig. 10.—Variation of tube parameters with frequency for helium-filled tube with 0·5% argon.

——— Reactance.
—·— Resistance.
– – – – Impedance.
— — Phase angle.

argon as possible. This confirms some of the earlier work by Bental and Benson,[21] who concluded that an increase in argon content reduces the tube impedance.

The impedance/frequency curves obtained were of the same form as those obtained by previous investigators,[21] the impedance increasing with frequency as predicted by the theory of van Geel[30] and Verhagen.[31] The results, however, differed somewhat from those obtained earlier,[21] which reported peaks in the curves at frequencies below 100 kc/s. It has already been mentioned, however, that the circuit previously used became unreliable at frequencies approaching 100 kc/s, probably because of the effects of stray capacitances. Yeh[19] has published results which also agree with the theory of van Geel and Verhagen. He has plotted reactance/resistance curves for various frequencies giving the loci of the impedance vectors, which are approximately circular as expected from the theory. These curves also demonstrate that the impedance vector reaches a maximum at a frequency just above that at which the reactance changes from inductive to capacitive. Yeh observed an approximate correlation between the transit time of ions in a glow-discharge tube and the frequency at which the tube impedance changes from inductive to capacitive. He states that if a small alternating voltage is superimposed on the d.c. discharge, ions produced at the edge of the negative glow will not reach the cathode in phase, and if the frequency is correct, will arrive exactly out of phase. According to Yeh, it is this frequency which determines the transition in the impedance. In calculating the ion transit time it is assumed that the ions are produced at the edge of the negative glow and that their velocity is a function of E/p (where E is the field strength and p the gas pressure) and therefore of position. The drift velocity is, in fact, assumed to be proportional to E/p at low fields and to $(E/p)^{\frac{1}{2}}$ at high fields. In his derivation, Yeh seems to have assumed an incorrect field distribution; it is not, as stated, the distribution suggested by Warren.[33, 34]

The calculation, which is given in the Appendix, indicates that the frequency at which zero reactance occurs is about 4·8 Mc/s for pure-helium-filled tubes and 6·3 Mc/s for pure-neon-filled tubes, thus giving peaks in the impedance/frequency curves just above these values respectively. The corresponding frequencies for the tubes with argon in the gas fillings could not be obtained, because reliable figures for the various tube constants involved are not available. More recent work by the authors on tubes having molybdenum cathodes, for which the tube constants are more accurately known, suggests that there is the correlation suggested by Yeh, although the experimental value occurs at a somewhat lower frequency than the calculated cross-over one. A further study is being made of this topic at present and the results will be reported in a future paper.

(4.6) Noise Characteristics

It has already been mentioned that internal oscillations in many tubes seriously limited the number of noise readings obtainable. In helium-filled tubes, however, only small random changes in the noise-voltage output occurred as the argon content changed. Internal oscillations have been observed previously[25] in special cerium-cathode tubes with different gas-fillings and also in certain commercial stabilizers with pure helium plus 2·2% argon gas-fillings when operated at large currents usually just higher than the upper specified normal working current. From the previous results, oscillations seemed to be associated with the amount of argon present, and this has now been shown to be so with the new specially-manufactured cerium tubes. For the present tubes, helium-filled types containing 2·5% or more argon exhibited oscillations. With 2·5% argon only one out of six tubes exhibited the effect; with 3·0% argon 50% of the tubes oscillated; and with 3·5% argon four out of six were found to oscillate. For all tubes exhibiting this phenomenon oscillations did not commence until the direct tube current had reached 20 mA. It is interesting to note that both helium- and neon-

filled tubes with no argon oscillated at all tube currents, and with all neon-filled samples oscillations were observed at all currents except those having 0·5% argon.

It has now been established that there is a correlation between the running voltage of a tube and the measured noise voltage at a given current. Tables 1 and 2 show this correlation for several

Table 1

RUNNING AND MEAN NOISE VOLTAGES FOR HELIUM-FILLED TUBES WITH DIFFERENT CATHODE MATERIALS RECORDED FOR TUBE CURRENTS OF 5 mA

Cathode material	Running voltage	Mean noise voltage
	volts	μV
Copper	177	259
Copper	159	196
Nickel	152	198
Copper	149	182
Molybdenum ..	122	144
Molybdenum ..	110	126
Cerium	111	120
Tantalum ..	105	118
Cerium	97	108
Zirconium ..	95	98
Zirconium ..	94	90

Table 2

RUNNING AND MEAN NOISE VOLTAGES FOR NEON-FILLED TUBES WITH DIFFERENT CATHODE MATERIALS RECORDED FOR TUBE CURRENTS OF 5 mA

Cathode material	Running voltage	Mean noise voltage
	volts	μV
Copper	150	232
Nickel	136	182
Nickel	131	178
Nickel	129	172
Nickel	115	148
Molybdenum ..	85	93
Tantalum ..	85	93
Cerium	72	70

experimental helium-filled and neon-filled tubes respectively for different cathode materials. This correlation explains why helium-filled tubes give more noise than neon-filled ones, with the same cathode material, which has been observed previously.[25] It also explains why gas-pressure variations from 30 to 50 mm Hg had little effect in determining noise characteristics, although cathode material is a very important parameter in this respect. The correlation is to be expected, because for most gases the probability of excitation and ionization by electron collision is a function of electron energy, so that since the tube noise is mostly due to collisions, it will also be a function of electron energy. Moreover, the majority of ionization collisions occur in the negative glow after the electrons have traversed the cathode-fall region and hence have an energy proportional to the cathode fall. It therefore follows that the tube noise is a function of the cathode fall or the running voltage approximately.

(5) ACKNOWLEDGMENTS

The work recorded has been carried out in the Department of Electrical Engineering at the University of Sheffield. The authors wish to thank Professor A. L. Cullen for facilities afforded in the laboratories of this Department and for his constant encouragement and advice; also Mr. P. M. Chalmers for carrying out many careful examinations of running-voltage/current and running-voltage/temperature characteristics and initial drifts of tubes. They also wish to acknowledge the kindness of The English Electric Valve Co. Ltd., in supplying all the special tubes for examination, for financial assistance with the work and for many helpful discussions and valuable suggestions.

(6) REFERENCES

(1) LAMMCHEN, K.: 'Die Zeitliche Konstanz von Glimmstrechken-Kennlinien und ihr Einfluss auf Spannungstabilisieningsschaltungen', *Hochfrequenztechnik und Electroakustik*, 1934, **44**, p. 193.

(2) KIRKPATRICK, G. M.: 'Characteristics of Certain Voltage-Regulator Tubes', *Proceedings of The Institute of Radio Engineers*, 1947, **35**, p. 485.

(3) TITTERTON, E. W.: 'Some Characteristics of Certain Voltage-Regulator Tubes', *Journal of Scientific Instruments*, 1949, **26**, p. 33.

(4) BENSON, F. A., CAIN, W. E., and CLUCAS, B. D.: 'Variations in the Characteristics of Some Glow-Discharge Voltage-Regulator Tubes', *ibid.*, p. 399.

(5) BENSON, F. A.: 'Initial Drifts in Running Voltage of Glow-Discharge Regulator Tubes', *ibid.*, 1950, **27**, p. 71.

(6) BENSON, F. A.: 'A Study of the Characteristics of Glow-Discharge Voltage-Regulator Tubes', *Electronic Engineering*, 1952, **24**, pp. 396 and 456.

(7) BENSON, F. A.: 'Variations in Extinction Voltages of Glow-Discharge Voltage-Regulator Tubes', *Journal of Scientific Instruments*, 1951, **28**, p. 186.

(8) BENSON, F. A.: 'Voltage Stabilization', *Electronic Engineering*, 1953, **25**, pp. 160 and 202.

(9) BENSON, F. A., and BACHE, H.: 'A Note on Temperature Coefficient of Running Voltage of Glow-Discharge Tubes', *Journal of Scientific Instruments*, 1952, **29**, p. 25.

(10) BENSON, F. A., and MAYO, G.: 'Effects of Ambient-Temperature Variations on Glow-Discharge Tube Characteristics', *ibid.*, 1954, **31**, p. 118.

(11) BENSON, F. A., and BENTAL, L. J.: 'Glow-Discharge Tubes', *Wireless Engineer*, 1956, **33**, p. 33.

(12) ANDREW, A. M.: 'The Design of Series-Parallel Voltage Stabilizers', *Electronic Engineering*, 1952, **24**, p. 385.

(13) HUNT, J. E. P.: 'Some Limitations in Voltage Stabilizers', *I.E.E. Students' Quarterly Journal*, 1952, **23**, p. 12.

(14) GROSZKOWSKI, J.: 'Lampa jarzeniowa jako indukcyjność', *Kwartalnik Telekomunikacyjny*, Styczen-Marzec, 1948, No. 1, p. 1.

(15) TOWNSEND, M. A., and DEPP, W. A.: 'Cold-Cathode Tubes for Transmission of Audio-Frequency Signals', *Bell System Technical Journal*, 1953, **32**, p. 1371.

(16) WILLIAMS, M. O.: 'Inertia Effects in Cold-Cathode Tubes', *Strowger Journal*, 1952, **8**, p. 106.

(17) 'The Impedance of Voltage-Stabilizer and Reference Tubes', *Mullard Technical Communications*, 1954, **1**, p. 219.

(18) HATTA, Y., SATO, T., and SAKATA, M.: 'The Measurement of the Low-Frequency Impedance of the Gaseous Electric Discharge', *Electrotechnical Papers*, Japan, 1952, **4**, p. 46.

(19) YEH, C.: 'Note on Positive-Ion Transit Time in Glow-Discharge Tubes', *Journal of Applied Physics*, 1956, **27**, p. 98.

(20) BENSON, F. A., and MAYO, G.: 'Impedance-Frequency Variations of Glow-Discharge Voltage-Regulator Tubes', *Electronic Engineering*, 1954, **26**, p. 206; and 'Impedance-Frequency Characteristics of some Glow-Discharge Tubes', *ibid.*, 1956, **28**, p. 124.

(21) BENSON, F. A., and BENTAL, L. J.: 'Glow-Discharge Stabilizers', *Wireless Engineer*, 1955, **32**, p. 330.
(22) BACHE, H., and BENSON, F. A.: 'Peak-Noise Characteristics of Glow-Discharge Voltage-Regulator Tubes', *Electronic Engineering*, 1952, **24**, p. 278.
(23) BACHE, H., and BENSON, F. A.: 'Peak-Noise Characteristics of Some Glow-Discharge Tubes', *Journal of Scientific Instruments*, 1953, **30**, p. 124.
(24) BACHE, H., and BENSON, F. A.: 'Mean Noise Characteristics of Glow-Discharge Voltage-Regulator Tubes', *Electronic Engineering*, 1952, **24**, p. 328.
(25) BENSON, F. A.: 'Gas-Filled Voltage Stabilizers', *Electronic and Radio Engineer*, 1957, **34**, p. 16.
(26) JURRIAANSE, T., PENNING, F. M., and MOUBIS, J. H. A.: 'The Normal Cathode Fall for Molybdenum and Zirconium in the Rare Gases', *Philips Research Reports*, 1946, **1**, p. 225.
(27) MEISSNER, J.: 'Über den Einflub der metastabilen Amegungazüstande auf die normale Stromdichte und den normalen Kathodenfall der Glimmentladung in Edelgasen un Edelgasgemischen', *Zeitschrift für Physik*, 1941, **117**, p. 325.
(28) EMELÉUS, K. G.: 'The Conduction of Electricity Through Gases' (Methuen, London, 1951), p. 86.
(29) JURRIAANSE, T.: 'The Influence of Gas Density and Temperature on the Normal Cathode Fall of a Gas Discharge in Rare Gases', *Philips Research Reports*, 1946, **1**, p. 407.
(30) VAN GEEL, C.: 'Untersuchungen von Gasentladungen mit Rücksicht auf ihre dynamischen Eigenschaften und ihre Stabilität', *Physica*, 1939, **6**, p. 806.
(31) VERHAGEN, C. J. D. M.: 'Impedanzmessungen an Gasentladungsröhren', *ibid.*, 1941, **8**, p. 361.
(32) HORNBECK, J. A.: 'The Drift Velocities of Molecular and Atomic Ions in Helium, Neon and Argon', *Physical Review*, 1951, **84**, p. 615.
(33) WARREN, R.: 'Field Measurements in Glow-Discharges with a Refined Electron Beam Probe and Automatic Recording', *ibid.*, 1955, **98**, p. 1650.
(34) WARREN, R.: 'Interpretation of Field Measurements in the Cathode Region of Glow-Discharges', *ibid.*, p. 1658.
(35) DRUYVESTEYN, M. J., and PENNING, F. M.: 'The Mechanism of Electrical Discharges in Gases of Low Pressure', *Reviews of Modern Physics*, 1940, **12**, p. 87.

(7) APPENDIX

The Transit Time of Ions Across the Cathode-Fall Space of a Glow-Discharge Tube

The transit time may be calculated from Maxwell's equations and a knowledge of the ion mobility in the cathode fall region.

Maxwell's equations give

$$\frac{dE}{dx} = \frac{d^2V}{dx^2} = 4\pi\rho \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The total charge density, ρ, is the difference between the positive-ion density, ρ_+, and the electron density, ρ_-, so that

$$\rho = \rho_+ - \rho_- = \frac{J_+}{v_+} - \frac{J_-}{v_-} \quad . \quad . \quad . \quad . \quad (2)$$

Let $$\gamma = J_-/J_+ \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and substitute this in eqn. (2):

$$\rho = J_+\left(\frac{1}{v_+} - \frac{\gamma}{v_-}\right) \quad . \quad . \quad . \quad . \quad (4)$$

where the total current density

$$J = J_+ + J_- = J_+(1+\gamma) \quad . \quad . \quad . \quad . \quad (5)$$

Eqn. (4) becomes

$$\rho = \frac{J}{1+\gamma}\left(\frac{1}{v_+} - \frac{\gamma}{v_-}\right) \quad . \quad . \quad . \quad . \quad (6)$$

and since $v_+ \ll v_-$ and $\gamma < 1$

$$\rho \simeq \frac{J}{1+\gamma}\left(\frac{1}{v_+}\right) \quad . \quad . \quad . \quad . \quad (7)$$

Substituting eqn. (7) in eqn. (1) gives

$$\frac{dE}{dx} \simeq \left(\frac{4\pi J}{1+\gamma}\right)\frac{1}{v_+} \quad . \quad . \quad . \quad . \quad (8)$$

Hornbeck[32] has carried out measurements on the drift velocity of ions in the cathode-fall region and has shown that for higher fields this velocity is given by

$$v_+ = k_2\left(\frac{E}{p}\right)^{1/2} \quad . \quad . \quad . \quad . \quad (9)$$

The substitution of eqn. (9) in eqn. (8) gives

$$\frac{dE}{dx} \simeq \frac{4\pi J}{(1+\gamma)k_2}\left(\frac{p}{E}\right)^{1/2} \quad . \quad . \quad . \quad . \quad (10)$$

If eqn. (10) is regarded as being an exact relationship,

$$E = \left[\frac{6\pi J p^{1/2} x}{(1+\gamma)k_2}\right]^{2/3} \quad . \quad . \quad . \quad . \quad (11)$$

for the boundary condition $E = 0$ at $x = 0$.

Warren[33, 34] has found that this 'mobility' field solution agreed very well with his practical results with the exception of a small region close to the negative glow.

If E_c is the field at the cathode,

$$E = E_c\left(\frac{x}{d_c}\right)^{2/3} \quad . \quad . \quad . \quad . \quad (12)$$

It follows from eqn. (9) that at the cathode the velocity v_c is given by

$$v_c = k_2(E_c/p)^{1/2} \quad . \quad . \quad . \quad . \quad (13)$$

so that $$v_+ = v_c(x/d_c)^{1/3} \quad . \quad . \quad . \quad . \quad (14)$$

Hence the transit time, τ_1, through this region, extending from $x = d$ to $x = d_c$, may be calculated from the expression

$$\tau_1 = \int_d^{d_c} \frac{dx}{v_+} = \frac{3}{2}\frac{d_c}{v_c}\left[1 - \left(\frac{d}{d_c}\right)^{2/3}\right] \quad . \quad (15)$$

In the low-field region near the negative glow the drift velocity, according to Hornbeck,[32] is given by

$$v_+ = k_1 E/p \quad . \quad . \quad . \quad . \quad . \quad (16)$$

It follows from eqn. (8) that, under such conditions and again with $E = 0$ when $x = 0$,

$$E = \left[\left(\frac{8\pi J}{1+\gamma}\right)\frac{p}{k_1}x\right]^{1/2} \quad . \quad . \quad . \quad . \quad (17)$$

If $E = E_1$ and $v_+ = v_d$ when $x = d$,

$$E = E_1(x/d)^{1/2} \quad . \quad . \quad . \quad . \quad (18)$$

and $$v_+ = v_d E/E_1 = v_d(x/d)^{1/2} \quad . \quad . \quad (19)$$

The transit time, τ_2, through this region is then given by

$$\tau_2 = \int_0^d \frac{dx}{v_d(x/d)^{1/2}} = \frac{2d}{v_d} \quad . \quad . \quad . \quad . \quad (20)$$

From eqn. (12), however,

$$E_1 = E_c\left(\frac{d}{d_c}\right)^{2/3} \quad . \quad . \quad . \quad . \quad . \quad (21)$$

so that $$v_d = k_1 E_1/p = k_1 E_c\left(\frac{d}{d_c}\right)^{2/3}/p \quad . \quad . \quad . \quad (22)$$

The normal cathode fall is given by

$$V_n = \int_0^d E_1\left(\frac{x}{d}\right)^{1/2} dx + \int_d^{d_c} E_c\left(\frac{x}{d_c}\right)^{2/3} dx \quad . \quad . \quad (23)$$

Therefore

$$V_n = \int_0^d E_c\left(\frac{d}{d_c}\right)^{2/3}\left(\frac{x}{d}\right)^{1/2} dx + \int_d^{d_c} E_c\left(\frac{x}{d_c}\right)^{2/3} dx \quad . \quad (24)$$

or $$V_n = \frac{E_c}{5}\left[3d_c + \frac{d^{5/3}}{3(d_c)^{2/3}}\right] \quad . \quad . \quad . \quad . \quad (25)$$

Thus $$V_n \simeq \frac{3}{5}E_c d_c \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

On substituting for v_c and E_c from eqns. (13) and (25) respectively in expression (15) for τ_1 it is found that

$$\tau_1 = \frac{3d_c p^{1/2}}{2\sqrt{(5)}k_2(V_n)^{1/2}}\left[3d_c + \frac{d^{5/3}}{3(d_c)^{2/3}}\right]^{1/2}\left[1 - \left(\frac{d}{d_c}\right)^{2/3}\right] \quad (27)$$

From eqns. (20), (22) and (25)

$$\tau_2 = \frac{2dp}{5k_1(d/d_c)^{2/3}V_n}\left[3d_c + \frac{d^{5/3}}{3(d_c)^{2/3}}\right] \quad . \quad . \quad (28)$$

The transit time across the cathode-fall region is therefore

$$\tau = \tau_1 + \tau_2$$
$$= \frac{3d_c p^{1/2}}{2\sqrt{(5)}k_2(V_n)^{1/2}}\left[3d_c + \frac{d^{5/3}}{3(d_c)^{2/3}}\right]^{1/2}\left[1 - \left(\frac{d}{d_c}\right)^{2/3}\right] + \frac{2dp}{5k_1(d/d_c)^{2/3}V_n}\left[3d_c + \frac{d^{5/3}}{3(d_c)^{2/3}}\right] \quad (29)$$

or $$\tau \simeq \frac{3\sqrt{(3)}(d_c)^{3/2}p^{1/2}}{2\sqrt{(5)}k_2(V_n)^{1/2}}\left[1 - \left(\frac{d}{d_c}\right)^{2/3}\right] + \frac{6d_c dp}{5k_1(d/d_c)^{2/3}V_n} \quad (30)$$

In the evaluation of the transit time the values of pd_c quoted by Druyvesteyn and Penning[35] were used. The transition value of E/p, where the field has been assumed to change from a low value to a high one, was obtained from the curves plotted by Hornbeck,[32] from which d/d_c is found from eqns. (21) and (26). In the case of helium-argon tubes the correct values of k_1 and k_2 were not known, so the figures for pure helium, obtained from Hornbeck's curves, have been used in calculating τ. Similarly, for neon-argon tubes the values of k_1 and k_2 for pure neon were used in the calculations. The frequency at which zero reactance occurs is then $1/2\tau$ if the correlation suggested by Yeh exists.

621.317.332.1.029.6

Paper No. 2396 R
Sept. 1957

AN AUTOMATIC SMITH DIAGRAM DISPLAY UNIT FOR USE AT LOW POWER LEVELS

By H. V. SHURMER, M.Sc., Ph.D., Associate Member.

(*The paper was first received 27th December, 1956, and in revised form 3rd May, 1957.*)

SUMMARY

An equipment is described whose primary purpose is to display variations in the r.f. admittance of low-level detector crystal valves during the adjustment process. For this purpose the input power level must not exceed about 5μW. The method is not, however, limited to this application but may be applied to microwave transmission systems generally.

Information relating to the r.f. admittance is derived from four detector probes mounted with spacings of $3\lambda/8$ along a standing-wave detector carriage. This is utilized electronically to actuate the deflecting plates of a cathode-ray tube in such a manner that a direct display of the Smith diagram is obtained on the face of the tube.

The above principles are incorporated in an equipment operating at 3·2 cm, but with an appropriate 4-probe standing-wave detector this could be used at any wavelength.

(1) INTRODUCTION

An earlier paper[1] describes an equipment in which the magnitude of the voltage standing-wave ratio (v.s.w.r.) of low-level detector crystal valves is directly measured on a meter. The present equipment gives a direct display on the face of a cathode-ray tube of the phase as well as the magnitude of the reflection coefficient at any desired waveguide reference plane. Satisfactory operation is obtained at power levels down to a few microwatts.

In the particular application for which the equipment has been developed an admittance transfer chart is used which is set over the face of the cathode-ray tube, enabling the Smith diagram to be referred to the plane of the semi-conducting contact within the crystal valve. The equipment may thus be used to give directly information regarding the point of contact whilst the adjustment of the valve is being made. Alternatively, it may be used for rapid admittance measurements on a batch of crystals. Used in conjunction with a suitable transfer chart the selection of several types of crystal valve which differ only in their specified admittance limits may be made in a single measurement, instead of by successive selection in several crystal holders.

(2) GENERAL BASIS OF OPERATION

A square-law crystal detector fed from a waveguide probe produces, under c.w. microwave excitation, a direct voltage output proportional to the square of the voltage across the waveguide, which is, of course, the vectorial sum of outgoing and reflected voltage wave components.

Consider four such probe units situated at spacings of $\lambda_g/8$ along the waveguide. With respect to the outgoing voltage vector, V_i, at each probe the reflected voltage vector will differ in phase by successive increments of $\pi/2$, and at the four probes is represented by V_{ra}, V_{rb}, V_{rc} and V_{rd}, respectively.

The four resultant voltage vectors are shown by V_a, V_b, V_c and V_d in Fig. 1 for arbitrarily chosen load conditions. The direct voltage output from each probe unit will be proportional to V_a^2, V_b^2, V_c^2 and V_d^2, respectively.

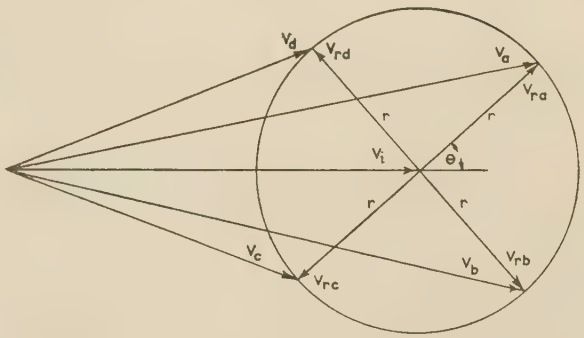


Fig. 1.—Vector diagram for 4-probe standing-wave detector.

If θ is the phase angle of the reflection coefficient at the plane of the first probe, it follows from elementary geometry that

$$V_a^2 = (V_i + r\cos\theta)^2 + (r\sin\theta)^2$$
$$V_c^2 = (V_i - r\cos\theta)^2 + (r\sin\theta)^2$$
$$V_a^2 - V_c^2 = 4V_i r\cos\theta$$

Similarly, $$V_b^2 - V_d^2 = 4V_i r\sin\theta$$

By taking the voltage difference between the outputs of alternate probe units, signals may thus be obtained which are proportional to $r\cos\theta$ and $r\sin\theta$. These two terms represent the abscissa and ordinate, respectively, of the reflection coefficient at the plane of the first probe. By applying direct voltages proportional to these terms to the X- and Y-plates of a cathode-ray tube, the reflection coefficient at this plane may be displayed as in the Smith diagram.

A spot may thus be produced on a Smith diagram grid placed over the face of the cathode-ray tube which represents the reflection coefficient at the plane of the first probe. If the system of four probes is mounted on the moving carriage of a conventional standing-wave detector, movement of the carriage enables the reference plane to be continuously variable, so that the diagram may be rotated about its origin.

In the next Section a particular equipment embodying these principles is described.

(3) APPLICATION OF PRINCIPLES

The first requirement in applying the above principles is to obtain four probe units which give identical outputs under the same input conditions. For this purpose four similar crystal detectors are selected, the output of each being fed into a potentiometer which allows for minor adjustments.

The output signals from the crystal detectors are of necessity small, making amplification essential. It is most convenient to provide square-wave modulation of the klystron power source at audio frequency, so that each output signal is an audio-frequency square-wave.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Shurmer is at the Research Laboratory of the British Thomson-Houston Co., Ltd.

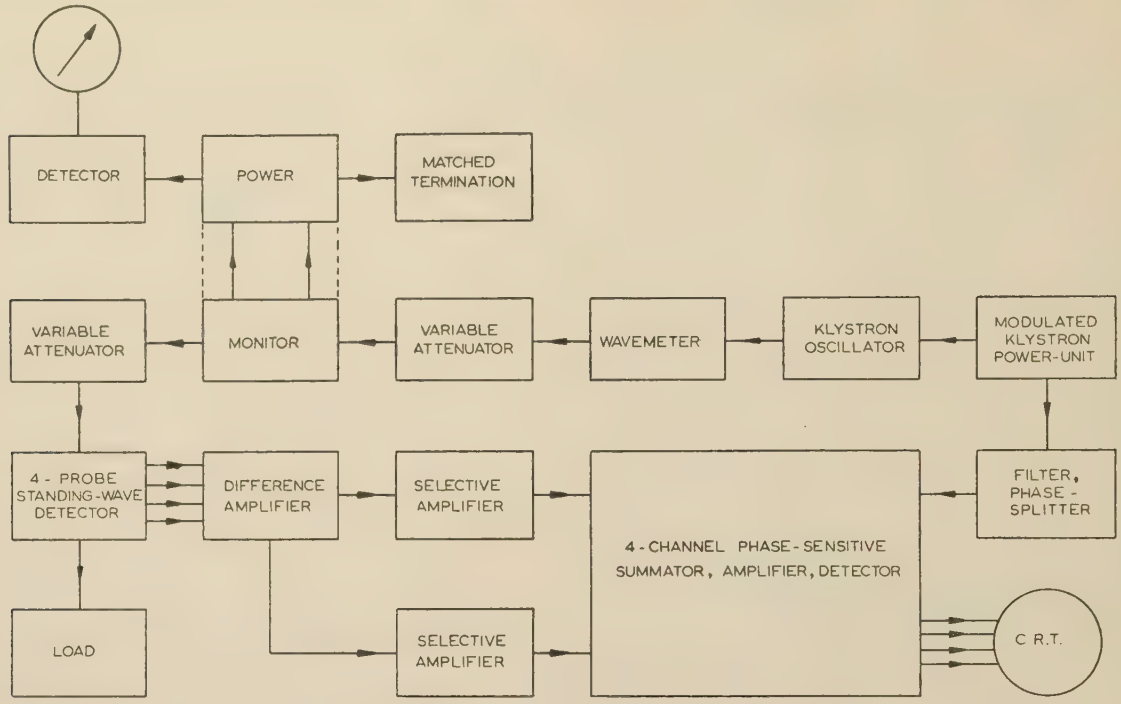


Fig. 2.—Block schematic of automatic Smith-diagram display unit.

The difference signal between alternate probe outputs is therefore also of audio frequency and is conveniently obtained by applying each pair of output signals to a difference amplifier. The two resultant difference signals obtained in this way are then amplified separately by high-gain selective audio amplifiers.

In order to preserve the sense of the difference signals each one is compared with a reference. The latter is conveniently derived from the square-wave modulator in the klystron power-supply unit.

In the present equipment two sinusoidal reference signals are used which are equal in magnitude but opposite in phase. These are obtained electronically by harmonic filtering and phase-splitting of the original reference signal.

Each difference signal is in phase with one reference signal and in anti-phase with the other. It is added to the two reference signals in separate stages. One of the resultant outputs increases with increasing difference signal, and the other decreases. The two output signals thus obtained are amplified and converted into direct voltages which are then applied to one pair of plates of a cathode-ray tube. The second pair of plates is similarly actuated by the other difference signal. This method allows the mean potential between the plates to remain constant, so that focusing is not affected by the deflecting voltages.

The above system results in deflecting potentials at the plates of the cathode-ray tube which are proportional to the difference signals from alternate probes in the waveguide. Hence, a direct display of the reflection coefficient at a variable plane in the waveguide is obtained.

(4) CIRCUIT DETAILS

A block schematic of the equipment is shown in Fig. 2.

(4.1) Four-Probe Standing-Wave Detector

The only special microwave component required in this equipment is a 4-probe standing-wave detector. It was shown in Section 2 that four probes are required spaced ideally at $\lambda/8$. Owing to the physical size of the probe units this spacing is impracticable. However, similar results may be obtained at a spacing of $3\lambda/8$ with a reduced broadband performance. As the equipment is required for measurements at a spot frequency only, this is no disadvantage and has therefore been adopted.

Crystal detectors of the CV2258 type are used in the probe units. At the very low power levels involved in this equipment all crystals of this type exhibit square-law behaviour. If the equipment were required to operate at a power level approaching 1 mW, attention would have to be paid to the square-law properties in the selection of suitable crystal detectors.

(4.2) Difference Amplifier

The signals derived from the four probe units on the standing-wave detector carriage are fed into four potentiometers, the setting of each being adjusted for equality of output under similar microwave conditions at the associated probe. The outputs from these potentiometers are taken in alternate pairs, each pair being fed into a double triode, connected as shown in Fig. 3.

The output is capacitance-coupled from one anode in each case. Such a circuit gives a resultant output at either anode proportional to the difference between the two input signals.

(4.3) Selective Amplifier

The selective amplifiers used in this application are standard units and have a narrow frequency pass band centred on the klystron modulation frequency. The output is taken from a low-impedance point in the amplification chain.

(4.4) Filter and Phase-Splitter

With reference to Fig. 4, the square-wave voltage derived from the modulator in the klystron power unit is filtered of its harmonic components by means of a twin-T stage, V_{1A}, as used in the selective amplifiers. This gives negative feedback across the triode at all but the selected frequency, which is sharply tuned.

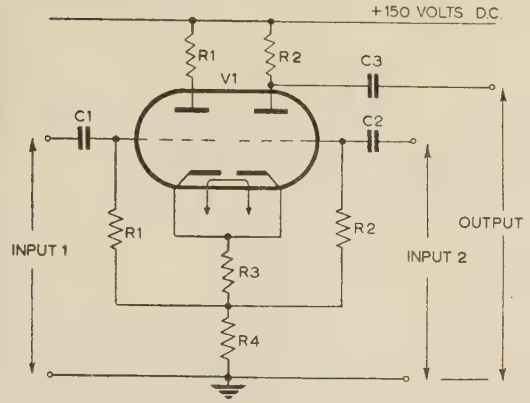


Fig. 3.—Triode connections in difference amplifier.

R1, 2 1 MΩ; R3 2·7 kΩ; R4 220 kΩ;
C1, 2, 3 0·01 μF; V1 12AT7.

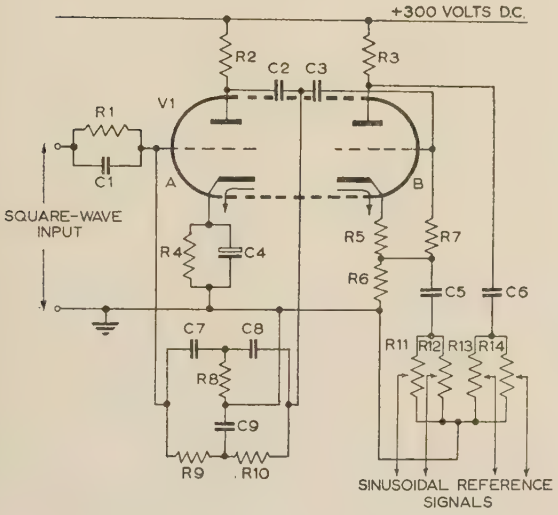


Fig. 4.—Filter and phase-splitter.

R1	680 kΩ	R7	1 MΩ	C1	100 pF
R2	200 kΩ	R8	75 kΩ	C2, 3	0·02 μF
R3, 6	47 kΩ	R9, 10	150 kΩ	C4	25 μF
R4	2·7 kΩ	R11, 12, 13, 14	0–1 MΩ	C5, 6	0·01 μF
R5	1·5 kΩ	V1	ECC83	C7, 8	330 pF
				C9	660 pF

The result is a sinusoidal output at modulation frequency of not more than about 5% harmonic distortion.

The output from the filter is coupled into the phase-splitter, for which the other half of the same valve, V_{1B}, is used. This has equal anode and cathode loads, output connections being made to each. This results in two signals of similar magnitude but of opposite phase. Each of these feeds two potentiometers, and the four resulting outputs are used as reference signals in the phase-sensitive summator.

(4.5) Phase-Sensitive Summator, Amplifier and Detector

The basic circuit for this part of the equipment is shown in Fig. 5. The output signal from each selective amplifier is fed into the one valve, V_1, of a pair of pentodes, the other valve, V_2, being fed from one of the two reference signals. The output from this stage is amplified by V_3 and feeds a full-wave rectifier, V_{4A} and V_{4B} and a voltage doubler V_{5A} and V_{5B} The resultant direct voltage is applied to one plate of a cathode-ray tube.

The speed with which the spot settles when the load conditions are changed is determined by the time-constant of the output

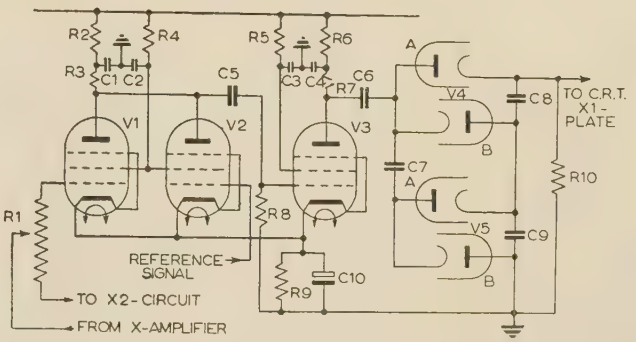


Fig. 5.—Phase-sensitive summator, amplifier, detector and voltage doubler.

R1	0–500 kΩ	R7	47 kΩ	C1, 2, 3, 4	8 μF
R2	15 kΩ	R8	100 kΩ	C5	0·002 μF
R3	1 kΩ	R9	120 Ω	C6, 7, 8, 9	0·1 μF
R4	56 kΩ	R10	10 MΩ	C10	50 μF
R5	120 kΩ	V1, 2, 3	6F12		
R6	10 kΩ	V4, 5	6D2		

circuit. This cannot be made too short or non-linearity is introduced by loading of the previous stage. In practice a delay of about one second is found satisfactory.

There are four circuits of the type described above, one for each combination of selective-amplifier output signal with a reference signal. The voltage doubler is necessary only in the case of the signals which are applied to the X-plates.

(4.6) Display

A 5in circular display has been found satisfactory. Facilities have been provided for inserting any desired form of diagram over the face of the tube so that the display may be interpreted directly in terms of the parameters at any selected plane in the microwave system.

(5) SETTING-UP PROCEDURE

The first requirement is to have the modulation frequency correctly set. This is done by tuning until either of the selective amplifiers gives maximum response as indicated on the amplifier output meter, which is normally short-circuited in the equipment but may be used for monitoring purposes.

The reference signals are then adjusted by means of four pre-set potentiometers. These are initially set to give zero output. The first one is then adjusted so as to deflect the spot on the cathode-ray tube by about two-thirds of the radial distance outwards. The potentiometer controlling the opposite plate is then employed to bring the spot back to the centre. This procedure is repeated with the other two potentiometers. The method ensures that for maximum deflecting voltage on any plate the signal voltage at the input to the phase-sensitive summator is always less than the reference voltage, and overloading is also avoided. Both these factors assist in maintaining linearity.

The rest of the setting-up procedure involves the signals derived from the standing-wave detector, and it is convenient to terminate the waveguide run in a short-circuit at this stage.

The four potentiometers associated with the difference amplifier are adjusted in pairs, using one selective amplifier to monitor each pair. With the four potentiometers initially set to give zero output the first one is adjusted for nearly maximum output. The maximum reading on the monitoring meter is noted as the standing-wave detector carriage is moved. The carriage is then set at a position in which the output reading is zero and the second potentiometer in the pair is adjusted to give the same maximum reading as before.

As the carriage is moved uniformly the output should now

vary between zero and a maximum twice as often as when the amplifier is fed from a single probe only. At the same time, the spot on the tube should follow either a horizontal or a vertical path, depending on which pair of plates is being supplied from the amplifier in question. The central position of the spot should correspond to zero output from the amplifier, and the positions of maximum deflection should correspond to maximum meter reading.

If the two limiting deflections differ slightly in magnitude they may be brought to equality by means of a balancing potentiometer connected at the input to the phase-sensitive summator. The gain of the selective amplifier is then adjusted so that the two limiting deflections lie on a diameter of the unit circle which contains the Smith diagram grid located over the face of the cathode-ray tube.

An exactly similar procedure is followed in setting up the circuit which feeds the other pair of plates of the tube. When this is done the spot should follow the unit circle as the position of the carriage is varied or, alternatively, as the short-circuit termination in the waveguide is moved.

If a variable attenuator is inserted in front of the short-circuit so that the load standing-wave ratio may be improved in stages, the familiar family of constant v.s.w.r. circles of the Smith diagram should be obtained on moving either the carriage or the short-circuit termination. The concentricity of these circles gives a check on the square-law properties of the probe units. Greater sensitivity in viewing this may be obtained by increasing the gain of both selective amplifiers.

(6) RANGES OF OPERATION

As in the expanded charts on which the central area only of the Smith diagram is plotted, so in this equipment ranges of expanded diagrams are obtainable. The expansion can be done very readily by removing fixed 10 dB steps of attenuation which are an integral part of the selective amplifiers.

The removal of 10 dB from each selective amplifier gives a diagram in which the limiting circle represents a v.s.w.r. of 0·52. The corresponding figure when 20 dB of attenuation is removed is 0·82. This represents the maximum sensitivity which is obtainable before noise troubles become serious.

(7) ACCURACY

The accuracy with which the reflection coefficient is displayed is limited only by the quality of the cathode-ray tube, provided that the following conditions are satisfied in the equipment:

(*a*) The probes are accurately spaced at $3\lambda/8$.
(*b*) The output from each probe obeys a square-law.
(*c*) The electronic circuits are free from distortion and stray pick-up.

Condition (*b*), which has been discussed in Section 4.1, should present no difficulty and condition (*c*) can be met by careful circuit design. The effects of departing from condition (*a*) are discussed below.

Any departure from centre frequency will produce an error in the displayed reflection coefficient, because the probe spacing will no longer be $3\lambda/8$. If this error is represented by a vector which is added to the true reflection coefficient, it may easily be shown that the maximum magnitude of this error vector, expressed as a fraction of the magnitude of the true reflection coefficient, is given approximately by twice the phase error, in radians, which results between adjacent probes in the waveguide.

This implies that the magnitude of the error vector can be 9·4% of that of the reflection coefficient for a 1% change in frequency. The phase of the error vector will be a function of the argument of the reflection coefficient.

The probe units have a Q-factor of approximately 30, which implies a reduction in the amplitude of the displayed reflection coefficient by some 26% for a 1% departure from centre frequency. Accurate use of this equipment is therefore restricted to a narrow bandwidth of a few megacycles on either side of the centre frequency which is set by the spacing of the probe units.

A comparison of experimental results has been made on a batch of 24 detector crystal valves of the CV2258 type, which were selected so that, with a particular holder, their admittance plot straddled the centre of the Smith diagram. Crystals outside specification, with a v.s.w.r. as low as 0·18, were included. Two readings of normalized admittance on each crystal using the automatic equipment were compared with two similar readings given by a conventional moving-carriage standing-wave detector, using in the latter case the mean of two readings of the position of the voltage minimum. Between readings the crystals were removed and reinserted in the holder. Results were expressed in terms of the normalized values of conductance and susceptance given by the Smith diagram.

It was found that the r.m.s. variations in G and B (the normalized conductance and susceptance) between the pairs of readings on the automatic equipment were 0·008 and 0·012, compared with 0·032 and 0·032, respectively, using the moving-carriage standing-wave detector. The r.m.s. variations between a set of measurements on the automatic equipment and on the moving-carriage standing-wave detector were 0·022 and 0·035 for G and B, respectively. It is thus seen that the discrepancy between readings obtained by the two methods is of the same order as the accuracy of the moving-carriage method.

(8) ACKNOWLEDGMENTS

This work was carried out under a contract placed by the Admiralty, to whom thanks are due for permission to publish. The author is indebted to his colleagues in the British Thomson-Houston Research Laboratory for encouragement and assistance in this work, particularly to Dr. I. M. Templeton.

(9) REFERENCE

(1) SHURMER, H. V.: 'A Direct-Reading Waveguide Standing-Wave Detector for Use at Low Power Levels', *Proceedings I.E.E.*, Monograph No. 119 R, January, 1955 (**102** C, p. 176).

621.396.677.75

Paper No. 2395 R
Sept. 1957

A METHOD OF ESTIMATING THE POWER RADIATED DIRECTLY AT THE FEED OF A DIELECTRIC-ROD AERIAL

By R. H. CLARKE, B.Sc.(Eng.).

(*The paper was first received 4th April, and in revised form 14th May*, 1957.)

SUMMARY

In a recent paper, Brown and Spector have shown that an end-fire aerial is a purely guiding structure, with radiation occurring only at the free end and at the feed end of the aerial. Such a system of radiation may be represented by a simple transmission-line analogue. The present paper illustrates this representation in the case of a waveguide-fed dielectric-rod aerial of uniform rectangular cross-section for use at 3 cm wavelengths and shows its application to finding the fraction of the total power supplied to the aerial which is radiated at the feed. This fraction was found to be 23% when the dielectric rod had a rectangular cross-section of 0·9 in × 0·4 in.

LIST OF SYMBOLS

Y_W = Characteristic admittance of the line representing the waveguide in the transmission-line analogue.
Y_D = Characteristic admittance of the line representing the dielectric rod.
G_A = Conductance representing the radiation from the free end of the dielectric rod.
G_B = Conductance representing the radiation from the junction between the waveguide feed and the dielectric rod.
Y_i = Input admittance of the section terminating the line representing the waveguide.
λ_0 = Free-space wavelength.
λ_W = Wavelength in the waveguide feed.
ρ = Reflection coefficient in the waveguide.
$|\rho|$ = Magnitude of the reflection coefficient.
ϵ_r = Relative permittivity of the dielectric forming the rod.
θ = Phase angle measured along the line representing the dielectric rod.

(1) INTRODUCTION

In a recent paper* Brown and Spector have shown that an end-fire aerial, and in particular the dielectric-rod, can be considered to support a non-radiating axial surface wave with radiation occurring at the discontinuities at the free end of the guiding structure and at the junction between this structure and the feed. It was pointed out that there was no direct experimental information about the relative amounts of power radiated at these two points in the case of the dielectric-rod aerial with a waveguide feed, and so a method has been devised to provide this information. The present paper describes this method and illustrates it by showing its application to the case of a rectangular dielectric-rod aerial of uniform cross-section for use at 3 cm wavelengths.

(2) THEORY OF THE METHOD

The mechanism of the radiation which Brown and Spector have shown to take place with a dielectric-rod aerial easily lends itself to a simple transmission-line circuit representation. Since the waveguide feed and dielectric rod, shown in Fig. 1, are purely guiding structures and no radiation occurs along their lengths, they may be represented by two transmission lines as

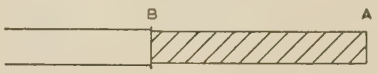


Fig. 1.—The dielectric-rod aerial.

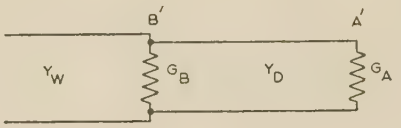


Fig. 2.—Transmission-line analogue representing the dielectric-rod aerial of Fig. 1.

shown in Fig. 2, the characteristic admittance of the line representing the waveguide being Y_W and that of the dielectric rod being Y_D. Radiation takes place only at the transition from the waveguide to the rod at B and at the transition from the rod to free-space at A. Such transitions may be represented by a shunt conductance and a series resistance, together with two lossless elements in a 4-terminal network.*

The nature of the radiation from A and B, however, suggests that the series resistance may be ignored and, as will be shown later, the experimental results confirm this assumption. Hence the radiation from A and B may be represented by the two conductances G_A and G_B. The changes in the wave configuration from the waveguide mode to the dipole mode on the rod and from that on the rod to the free-space plane wave suggest that there may be changes of phase at both A and B. (Indeed these phase changes are represented by the lossless elements in the 4-terminal network mentioned above.) Thus the electrical length A′B′ will differ from the actual length AB of the rod, but this difference will not affect the analysis, which is in terms of the change in electrical length only.

Now if, instead of terminating abruptly, the dielectric rod is tapered at the free end, as is usually done in practice to minimize reflections from this end, then the rod is matched and the conductance G_A will have a value Y_D. In this case, since the voltages applied to G_B and Y_D have the same magnitude, the ratio

$$\frac{\text{Power radiated at the feed}}{\text{Power radiated at the free end}} = \frac{G_B}{Y_D} \quad . \quad . \quad . \quad (1)$$

and the percentage of the total power supplied which is radiated directly at the waveguide feed is given by the expression

$$\text{Percentage of directly radiated power} = \frac{100\,G_B}{Y_D + G_B} \quad . \quad (2)$$

* BROWN, J., and SPECTOR, J. O.: 'The Radiating Properties of End-Fire Aerials', *Proceedings I.E.E.*, Paper No. 2216 R, January, 1957 (104 B, p. 27).

* FELSEN, L. B., and OLINER, A. A.: 'Determination of the Equivalent-Circuit Parameters for Dissipative Microwave Structures', *Proceedings of the Institute of Radio Engineers*, 1954, 42, p. 477.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Clarke is in the Electrical Engineering Department, University College, University of London.

In eqns. (1) and (2) it is assumed that the wave admittance, Y_D, is real.

Thus it is required to know the values of the conductances and admittances in the equivalent circuit of the aerial. These may be related to the reflection coefficient in the waveguide, and since the modulus of the reflection coefficient can be obtained from the standing-wave pattern in the guide the values of the conductances and admittances may be deduced. A further standing-wave pattern may be obtained by placing a brass reflector at A, Fig. 1, flush with the plane of cross-section; the conductance G_A thus becomes infinite.

(3) EXPERIMENTAL INVESTIGATION

A rectangular waveguide with internal dimensions 0·9 in × 0·4 in was used to feed a polythene dielectric rod of uniform cross-section, as shown in Fig. 1, and of length 7 in, i.e. about 6 λ_0. This length was chosen in order that the number of wavelengths accommodated by the rod at the extremes of the frequency range of the klystron oscillator should differ by about unity; this length was also sufficient to ensure that the surface wave would be efficiently established over the last few wavelengths on the rod. The plane brass reflector was about 6 in square. Although the surface wave is theoretically of infinite extent, the dimensions of the rod were large enough compared with the wavelength for the bulk of the energy to travel within a few wavelengths of the rod; such a reflector has sufficient area to ensure almost perfect reflection. The end of the dielectric rod inserted into the waveguide was tapered to reduce reflections at this point. By applying Feenberg's method* in a separate test it was found that in the frequency range considered the reflection coefficient of the taper was less than 0·02 and could thus be ignored.

Fig. 3 shows how the magnitude of the reflection coefficient

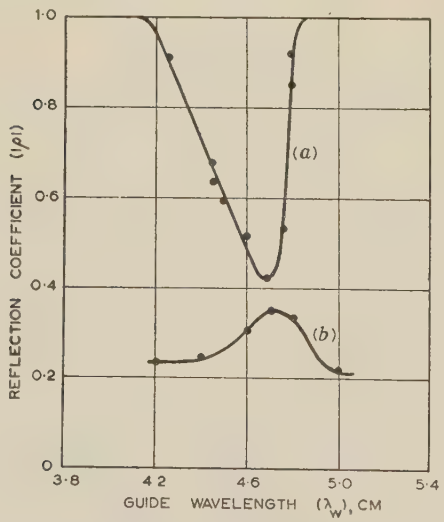


Fig. 3.—Observed variation of the magnitude of the reflection coefficient with guide wavelength.

(*a*) With reflector ($|\rho|_{max} = 1$; $|\rho|_{min} = 0{\cdot}42$).
(*b*) Without reflector ($|\rho|_{max} = 0{\cdot}35$; $|\rho|_{min} = 0{\cdot}23$).

varies with the wavelength, λ_W, in the waveguide, both quantities being deduced from standing-wave patterns in the waveguide obtained at a number of frequencies within the range of the klystron oscillator.

* BARLOW, H. E. M., and CULLEN, A. L.: 'Microwave Measurements' (Constable, 1950) Section 6.5.

(4) EXAMINATION OF THE RESULTS

The behaviour of the aerial with varying frequency, shown in Fig. 3, is examined with reference to the transmission-line analogue shown in Fig. 2. This examination does not yield a unique value for the required percentage of directly radiated power, and a physical approach to the problem has therefore been adopted in order to resolve certain ambiguities. However, a more rigorous proof of the validity of some of the results will be found in Section 7. It must be noted that the characteristic admittances and conductances are assumed to remain constant with varying frequency: this is not a valid assumption, but since the values used in the calculations are always given as ratios of the characteristic admittance of the waveguide the errors introduced by this assumption are unlikely to be very great.

As the frequency varies the electrical length of the line A′B′ varies. With the reflector in position the conductance G_A becomes infinite, so that when the line is an integral number of half-wavelengths long the transferred admittance of the line at B′ is also infinite. In this case the magnitude of the reflection coefficient in the waveguide is unity, as the results show. The fact that the reflection coefficient can have a value of unity shows that the assumption made in Section 2, namely that the series-resistance term in the equivalent circuit is negligible, is a valid one, since if the series-resistance term were appreciable then the reflection coefficient could never be unity. The reflection coefficient is a minimum when the transferred admittance is least, i.e. when the line A′B′ is an odd number of quarter-wavelengths long. In this case the line appears as an open-circuit and

$$|\rho|_{min} = \frac{|Y_W - G_B|}{Y_W + G_B} \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Substituting the value $|\rho|_{min} = 0{\cdot}42$ yields

$$\frac{G_B}{Y_W} = 0{\cdot}41 \quad \text{if } G_B < Y_W \quad . \quad . \quad . \quad . \quad (4a)$$

or

$$\frac{G_B}{Y_W} = 2{\cdot}45 \quad \text{if } G_B > Y_W \quad . \quad . \quad . \quad . \quad (4b)$$

The value given by eqn. (4*a*) appears to be the correct one, since if $G_B = 0$ there is no direct radiation; this will be confirmed later.

The maximum value of the reflection coefficient without the reflector occurs, as the curves show, at almost the same frequency as the minimum value of the reflection coefficient with the reflector, i.e. when the line A′B′ is an odd number of quarter-wavelengths long, in which case the transferred admittance at B′ is Y_D^2/G_A and

$$|\rho|_{max} = \frac{|Y_W - (G_B + Y_D^2/G_A)|}{Y_W + (G_B + Y_D^2/G_A)} \quad . \quad . \quad . \quad (5)$$

Substituting the value $|\rho|_{max} = 0{\cdot}35$ in this expression gives

$$\frac{G_B}{Y_W} + \frac{Y_D^2}{G_A Y_W} = 0{\cdot}48 \quad \text{if } \left(G_B + \frac{Y_D^2}{G_A}\right) < Y_W \quad . \quad (6a)$$

or

$$\frac{G_B}{Y_W} + \frac{Y_D^2}{G_A Y_W} = 2{\cdot}08 \quad \text{if } \left(G_B + \frac{Y_D^2}{G_A}\right) > Y_W \quad . \quad (6b)$$

Since both terms on the left-hand side of eqns. (6*a*) and (6*b*) are positive, the solution given by eqn. (4*b*) is impossible; hence

$$\frac{G_B}{Y_W} = 0{\cdot}41 \quad . \quad . \quad . \quad . \quad . \quad . \quad (4a)$$

This value, when substituted into eqns. (6*a*) and (6*b*), gives

$$\frac{Y_D^2}{G_A Y_W} = 0{\cdot}07 \quad . \quad . \quad . \quad . \quad . \quad (7a)$$

or

$$\frac{Y_D^2}{G_A Y_W} = 1{\cdot}67 \quad . \quad . \quad . \quad . \quad . \quad (7b)$$

The minimum value of reflection coefficient without the reflector occurs when the line A′B′ is an integral number of half-wavelengths long, for which the transferred admittance is G_A and

$$|\rho|_{min} = \frac{|Y_W - (G_A + G_B)|}{Y_W + (G_A + G_B)} \quad . \quad . \quad . \quad . \quad (8)$$

Substituting the value $|\rho|_{min} = 0{\cdot}23$ and eqn. (4*a*) in eqn. (8) yields

$$\frac{G_A}{Y_W} = 0{\cdot}23 \quad \text{if} \quad (G_A + G_B) < Y_W \quad . \quad . \quad (9a)$$

or

$$\frac{G_A}{Y_W} = 1{\cdot}17 \quad \text{if} \quad (G_A + G_B) > Y_W \quad . \quad . \quad (9b)$$

In order to make the necessary substitutions in eqn. (2) to find the percentage of directly radiated power it is necessary to choose between the alternative values presented by eqns. (7*a*) and (7*b*) and by eqns. (9*a*) and (9*b*). This is achieved in the case of eqns. (9*a*) and (9*b*) by noting that G_A is likely to be of the same order as Y_0, the admittance of free space. Now $Y_0/Y_W = \lambda_W/\lambda_0$, which is greater than unity, and hence it seems likely that $G_A/Y_W > 1$, which leads to the choice of eqn. (*b*) as being correct in preference to (9*a*). Substituting eqn. (9*b*) in eqns. (7*a*) and (7*b*) gives

$$\frac{Y_D}{Y_W} = 0{\cdot}29 \quad . \quad . \quad . \quad . \quad . \quad . \quad (10a)$$

or

$$\frac{Y_D}{Y_W} = 1{\cdot}4 \quad . \quad . \quad . \quad . \quad . \quad . \quad (10b)$$

Since Y_D corresponds to propagation along the dielectric rod, then

$$Y_0 < Y_D < Y_0\sqrt{\epsilon_r} \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where ϵ_r is the relative permittivity of the dielectric, and in the case of polythene is about 2·56. Now $Y_0 = Y_W(\lambda_W/\lambda_0) \simeq \frac{4}{3}Y_W$, and hence the inequality (11) becomes

$$\frac{4}{3} < \frac{Y_D}{Y_W} < 2 \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This shows that eqn. (10*b*) is the correct one. Substituting eqns. (9*b*) and (10*b*) into eqn. (2) gives the percentage of the total power supplied to the aerial which is radiated at the feed as 23%.

(5) DISCUSSION OF THE RESULTS AND CONCLUSIONS

There seems to be no information available that affords a direct confirmation of this estimate of 23% as the amount of radiation which is radiated directly at the waveguide feed, but the value obtained is not inconsistent with the facts that are to hand. In their work Brown and Spector assumed that, in the case of a circular dielectric rod of about half a free-space wavelength in diameter, 6% of the total power is radiated directly, and subsequently confirmed this figure indirectly by showing that it led to theoretical results which were consistent with those obtained in practice. The dielectric rod used in the present work, however, was of rectangular cross-section with a narrow dimension (which may be considered as the critical dimension) of about a quarter of a free-space wavelength, i.e. half the diameter for which 6% direct radiation is obtained. Now the larger the cross-sectional dimensions of the rod the more efficiently the surface wave will be launched, since for a rod of infinite dimensions there will be no direct radiation while for a rod of vanishingly small cross-section all the radiation will be direct. On this reasoning the value of 23% is quite acceptable.

It is not possible to make a precise estimate of the accuracy of this result, but since in all the stages of the calculations the normalized admittances are in good agreement with the estimates of these values based on physical arguments it does not appear unreasonable to suppose the result to be accurate to one part in five at least.

Further, since this apparently consistent result was obtained by representing the aerial by the simple transmission-line analogue shown in Fig. 2, it would seem that such a representation is a valid one, at least for certain applications.

(6) ACKNOWLEDGMENTS

The author wishes to record his thanks to Prof. H. E. M. Barlow for the use of the laboratory, workshops and equipment, and to Dr. J. Brown for his guidance and encouragement.

(7) APPENDIX: STATIONARY VALUES OF THE MAGNITUDE OF THE REFLECTION COEFFICIENT

Referring to Fig. 2, with the reflector in position, G_A becomes infinite. Let θ be the change in phase angle (which is frequency-dependent) along the line A′B′, then the admittance terminating the line Y_W is given by

$$Y_i = G_B - jY_D \cot\theta \quad . \quad . \quad . \quad . \quad (13)$$

Hence, assuming $G_B < Y_W$, the reflection coefficient is

$$\rho = \frac{Y_W - (G_B - jY_D \cot\theta)}{Y_W + (G_B - jY_D \cot\theta)} \quad . \quad . \quad . \quad (14)$$

and

$$|\rho|^2 = \frac{(Y_W - G_B)^2 + Y_D^2 \cot^2\theta}{(Y_W + G_B)^2 + Y_D^2 \cot^2\theta} \quad . \quad . \quad . \quad (15)$$

Differentiating the right-hand side of eqn. (15) with respect to θ and equating to zero to obtain the turning values shows that these occur at values of θ of $n\pi$ and $(n+1)\pi/2$, where n is an integer. When $\theta = n\pi$ the line A′B′ is an integral number of half-wavelengths long and $|\rho|$ is a maximum, i.e.

$$|\rho|_{max} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (16a)$$

The minimum value of $|\rho|$ occurs when $\theta = (n+1)\pi/2$, i.e. when the line A′B′ is an odd number of quarter-wavelengths long, in which case

$$|\rho|_{min} = \frac{Y_W - G_B}{Y_W + G_B} \quad . \quad . \quad . \quad . \quad (17a)$$

If, however, $G_B > Y_W$, then

$$|\rho|_{max} = 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (16b)$$

and

$$|\rho|_{min} = \frac{G_B - Y_W}{G_B + Y_W} \quad . \quad . \quad . \quad . \quad (17b)$$

Without the reflector in position the admittance terminating the line, Y_i, is given by

$$Y_i = G_B + Y_D \frac{G_A + jY_D \tan\theta}{Y_D + jG_A \tan\theta} \quad . \quad . \quad . \quad (18)$$

From this the reflection coefficient is found to be

$$\rho = \frac{Y_D(Y_W - G_A - G_B) + j(Y_W G_A - G_A G_B - Y_D^2)\tan\theta}{Y_D(Y_W + G_A + G_B) + j(Y_W G_A + G_A G_B + Y_D^2)\tan\theta} \quad . \quad . \quad . \quad . \quad (19)$$

By an analysis similar to that given above it may be shown that when $\theta = n\pi$,

$$|\rho| = \frac{|Y_W - (G_A + G_B)|}{Y_W + (G_A + G_B)} \quad . \quad . \quad . \quad . \quad (20)$$

and that when $\theta = (n + 1)\pi/2$,

$$|\rho| = \frac{|Y_W - (G_B + Y_D^2/G_A)|}{Y_W + (G_B + Y_D^2/G_A)} \quad . \quad . \quad . \quad (21)$$

621.316.842:621.315.55

Paper No. 2398 M
Sept. 1957

THE ELECTRICAL CHARACTERISTICS OF A NICKEL-CHROMIUM-ALUMINIUM-COPPER RESISTANCE WIRE

By C. DEAN STARR, Ph.D., and T. P. WANG, Ph.D.

(*The paper was first received* 15*th April, and in revised form* 3*rd June,* 1957.)

SUMMARY

The changes in electrical resistance and temperature coefficient of resistance during the heat treatment of a nickel-chromium-aluminium-copper alloy, known commercially as Evanohm, are presented. From the nature of the resistance/temperature curve and X-ray diffraction data, it is proposed that the electrical changes be attributed to short-range order.

Data are presented to show that both the mean temperature coefficient of resistance and the minor deviations in linearity in the resistance/temperature curve vary in a systematic manner during heat treatment and that the changes are independent of prior thermal history.

Resistance changes after heat treatment, caused by either surface tarnishing, oxidation, or the application of strain during fabrication, are discussed.

(1) INTRODUCTION

The resistance characteristics of a nickel-chromium-aluminium-copper resistance alloy, known commercially as Evanohm, were recently published.[1] It was shown that Evanohm has many desirable properties as a precision resistance alloy. For example, its stability of resistance up to 120° C is excellent, and may be superior even to Manganin for precision high-resistance devices. It possesses the additional advantage that its resistance is constant over a considerable temperature range, compared with Manganin. Finally, its high resistivity (800 ohms per circular-mil-foot), high yield and tensile strength, and ductility make it very suitable for obtaining fine wire for the production of high resistance per unit length.

The high resistivity and low temperature coefficient of resistance of this alloy are obtained by suitable heat treatment. As a part of his investigation, Arnold[1] heat-treated Evanohm by passing an electric current through an 0·004 8 in diameter wire. He observed that the mean temperature coefficient was a direct function of the time of heat treatment, but the minor variation in linearity of the temperature coefficient curve was dependent on prior thermal history. Previous tests at the authors' laboratory did not corroborate this observation. The purpose of the paper, therefore, is primarily to present additional data to show the resistance characteristics of Evanohm after different heat treatments. As a corollary, it is desirable to amplify and corroborate some of Arnold's observations.

(1.1) Nature of the Resistance Change During Heat Treatment

Evanohm is a metastable solid-solution alloy that has a nominal composition of 75% Ni, 20% Cr, 2·75% Al and 2·00% Cu. The resistance at room temperature and temperature coefficient of resistance (20–100° C), after heat treatment above about 600° C, are almost independent of the heat-treating temperature. The resistance at temperature (hot resistance), however, increases with increasing temperature. Below about 600° C, the room-temperature resistance increases and the temperature coefficient of resistance (20–100° C) decreases with decreasing heat-treatment temperature. This change is most pronounced at temperatures between 600 and 400° C. The magnitude of the change at any temperature is a function of the time of heat treatment. The absence of any detectable precipitation during the annealing and subsequent heat-treating cycle, in which the resistance can be made to increase by as much as 16%, suggests that the resistance change is caused by an order-disorder reaction. In this alloy, as in the high-content nickel-chromium alloys such as 80% Ni, 20% Cr, the resistance change occurs over a range of temperatures rather than at a specific temperature, which is indicative that the ordering is of a short- rather than long-range type.[2] This is also consistent with X-ray investigations, in which no superstructure lines are visible even when manganese K_α-radiation is used to overcome the small difference in atomic scattering factors between nickel and chromium.[3] Ordering can be detected, however, either by measuring heat evolution or lattice contraction. While measurements of specific heat have not been made, lattice contractions of Evanohm have been calculated after various stages of heat treatment by measuring the decrease in spacing of the [220] plane using chromium K_α-radiation. The data from this high-angle reflection line, shown in Table 1, in conjunction with the absence of a discontinuity in the equilibrium resistance curve of Evanohm as a function of temperature, are consistent with changes attributed to short-range order.

(1.2) Heat Treatment of Evanohm

The effect of time and temperature of heat treatment on the electrical properties of Evanohm have been extensively investigated. To minimize experimental errors, the following procedure has been used. The sample mount consists of a ½ in-diameter ceramic bobbin containing a spiral groove. Near each end of the bobbin a 0·125 in Tophet A (80% Ni, 20% Cr) lead is attached. A very small bead of silver solder is then melted on each Tophet A lead, and the resistance wire is soldered to the leads. By inserting an end of the wire sample into the solder just prior to its solidification, the amount of heat treatment occurring during sample preparation is negligible. By using fairly deep grooves on the bobbin, very loose coils also can be prepared, which eliminates stresses on the wire.

Heat treatment is accomplished by inserting the sample in the hot zone of a laboratory tube furnace. At the end of the cycle, the sample is rapidly withdrawn and partially cooled in the end of the tube in a cracked ammonia atmosphere (75% H_2, 25% N_2) before exposing the resistance wire to air. Temperature coefficients are then obtained on the samples heat treated *in situ* by alternately placing the assembly in a 20° C and 100° C oil bath. Since the sample preparation precedes heat treatment, the spurious results arising from heat treatment and the introduction of stresses during sample preparation are eliminated completely. The only precaution necessary is to prevent hydrogen or cracked ammonia from burning in the furnace during heat treatment. Such a process would tend to cause the wire to be partially annealed, resulting in a high temperature coefficient. However, it is also possible to heat-treat a sample again and subsequently lower its temperature coefficient.

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Starr and Dr. Wang are with Wilbur B. Driver Co., New Jersey, United States.

Table 1

EFFECT OF HEAT TREATMENT ON THE DECREASE IN LATTICE SPACING OF EVANOHM

Heat treatment	Resistivity	γ (20–100° C)	d_{220}*	d_{av}	a_{0av}	$\Delta a_0/a_0$
	ohms per circular-mil-foot	part in 10^6 per deg C	Å	Å	Å	%
Strand anneal	729	+53	1·2609 1·2615	1·2611	3·5669	0
Strand anneal + 565° C 20 min	788	+ 1	1·2604 1·2608	1·2606	3·5655	−0·04
Strand anneal + 565° C 67 hours	801	+36	1·2596 1·2600	1·2599	3·5635	−0·10

(XRD camera, chromium K_α radiation, 35 kV (peak), 17 mA, 2-hour exposure.)

* d_{220} = Range of d-values.
d_{av} = Results of ten measurements on [220] doublet.

A typical curve of the increase in both hot resistance and resistance at room temperature, and the change in temperature coefficient, is shown in Fig. 1. Initially, the resistance increases

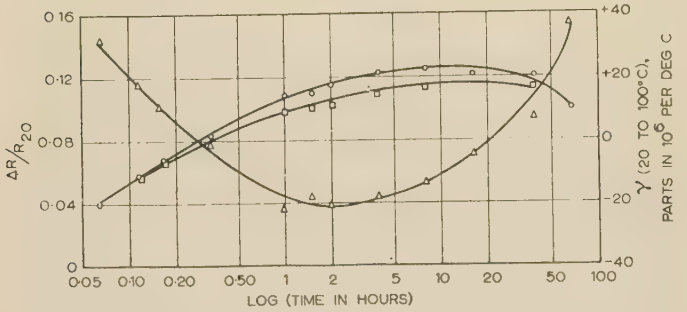


Fig. 1.—Effect of time on the change in resistance and temperature coefficient of Evanohm heat-treated at 565° C. (Initial values of ρ and γ are 733 ohms per circular-mil-foot and +54 parts in 10^6 per deg C.)

○ = Room-temperature resistance.
⊡ = Hot resistance.
△ = γ (at 20–100° C).
$\Delta R/R$ = Change in resistance after heat treatment at the time indicated.

and the temperature coefficient decreases. In production, the cycle is so chosen that the time of heat treatment corresponds to the point at which the temperature-coefficient curve first intersects the zero axis. This is referred to as the 'first zero temperature coefficient'. Continued heating beyond this point causes the temperature-coefficient curve to become more negative while the resistance continually increases. Eventually the temperature coefficient reaches its most negative value and thereafter increases and intersects the zero axis again. This second intersection is termed the 'second zero temperature coefficient'. The alloy in this state has a higher resistivity and a resistance/temperature curve that is not as linear as an alloy heat-treated to the first zero temperature coefficient.

(1.3) Relation between Resistivity and Temperature Coefficient of Resistance

The correlation of the data on resistivity and temperature coefficient of resistance (20–100° C) at a series of temperatures is shown in Fig. 2. This curve reveals that, during initial heat treatment, a unique value of the temperature coefficient for each value of resistivity, independent of prior thermal treatment, is obtained. This relation can be expressed as follows:

$$\gamma = -a(\rho - \rho_0) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

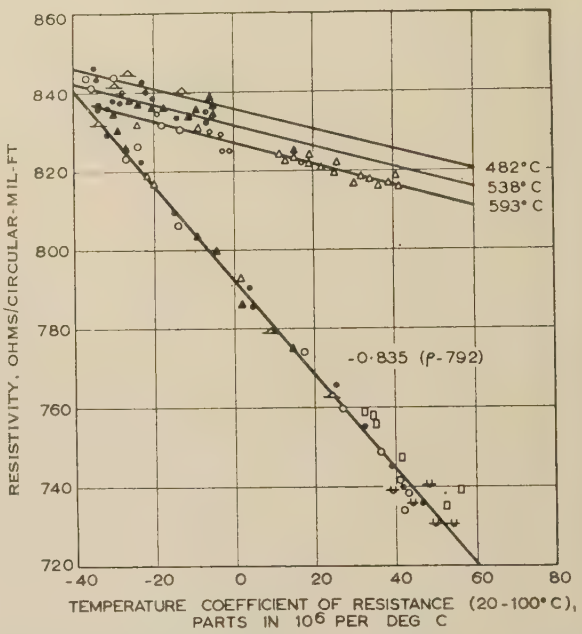


Fig. 2.—Relation between resistivity and temperature coefficient of Evanohm after heat treatment at the designated temperatures.

Type of sample	Temperature of heat treatment
○ Strand annealed	593° C
△ Hard drawn	593° C
● Strand annealed	538° C
▲ Hard drawn	538° C
⊙ Strand annealed	482° C
△ Hard drawn	482° C
⊎ Strand annealed	None
□ Strand annealed	None

where γ = Temperature coefficient of resistance, parts per million (20–100° C).
a = Constant.
ρ = Resistivity, ohms per circular-mil-foot.
ρ_0 = Resistivity at zero value of γ.

Eqn. (1) can be used to predict the course of heat treatment. For two tests on the same wire, eqn. (1) can be written as follows:

$$\frac{\rho_2}{\rho_1} = 1 + \frac{l}{a\rho_1}(\gamma_1 - \gamma_2) \quad . \quad . \quad . \quad . \quad . \quad (2)$$

or in terms of resistance,

$$\frac{R_2}{R_1} = 1 + \frac{l}{aR_1d^2}(\gamma_1 - \gamma_2) \quad . \quad . \quad . \quad . \quad (3)$$

where R_2 = Resistance corresponding to γ_2.
R_1 = Resistance corresponding to γ_1.
l = Length of wire.
d = Diameter of wire.

After one measurement when R_1 is set, the constant l/aR_1d^2 is fixed, and the relationship is valid for predicting the course of heat treatment. Eqn. (3) is essentially the one observed in Reference 1; the constant for the test conditions was 2000.

(2) TEMPERATURE COEFFICIENT OF ELECTRICAL RESISTANCE OF EVANOHM

The change in resistance between −60° C and +100° C is essentially linear. This is the basis for using the mean temperature coefficients referred to above. There are, however, minor deviations from linearity. It has been shown that the resistance/temperature relation, between 20° and 300° C, can be represented very closely by the following parabolic law:[1]

$$R_T = R_{20}[1 + \alpha(T - 20) + \beta(T - 20)^2] \quad . \quad . \quad (4)$$

where R_T = Resistance at a temperature T.

R_{20} = Resistance at a temperature of 20° C.

α and β = Constant coefficients.

The deviation from linearity is reflected in the coefficient β in eqn. (4). If β is zero, the resistance change is linear with temperature and the temperature coefficient of the alloy is α. The data in Table 2 were obtained by heat-treating strand-annealed Evanohm on a ceramic bobbin for various times in a cracked ammonia atmosphere at 565° C, as outlined previously. A new sample was used for each test, except the last datum in Table 2. That particular sample was cumulatively heat-treated to trace out the temperature-coefficient/time curve. The data in Table 2 show that, between −60° and +100° C, α falls systematically from $+57 \times 10^{-6}$ for the annealed wire to a negative maximum of -16×10^{-6}, and thereafter changes again in a positive direction. This is in agreement with previous investigations.[1] The variations in β, unlike those reported previously, also show a very systematic pattern, in that it changes from $-0{\cdot}051 \times 10^{-6}$ to $-0{\cdot}025 \times 10^{-6}$ and thereafter increases consistently in a negative direction. The difference in the magnitude of β arises because these values were determined over the temperature range −60–100° C instead of the range 20–300° C used by Arnold. As shown in Figs. 3–5 and Table 2, the portion of the curves from 20 to 100° C are almost linear, β being about $-0{\cdot}030 \times 10^{-6}$, which is of the order of magnitude previously observed.

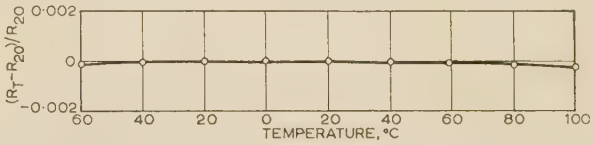


Fig. 3.—Change in resistance of Evanohm between −60 and +100° C.
(Evanohm was heat-treated to about the first zero temperature coefficient: 565° C, 20 min.)
$R_T = R_{20}[1 + 0(T - 20) + (-0{\cdot}031 \times 10^{-6})(T - 20)^2]$

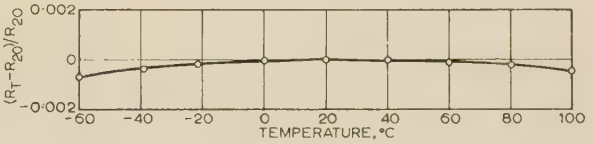


Fig. 4.—Change in resistance of Evanohm between −60 and +100° C.
(Evanohm was heat-treated to about the second zero temperature coefficient: 565° C, 27 hours.)
$R_T = R_{20}[1 + 2 \times 10^{-6}(T - 20) - (0{\cdot}081 \times 10^{-6})(T - 20)^2]$

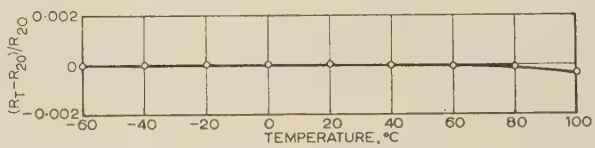


Fig. 5.—Change in resistance of Evanohm between −60 and 100° C.
(Evanohm was heat-treated in production to about the first zero temperature coefficient: 504° C, 1 hour.)
$R_T = R_{20}[1 + (-2 \times 10^{-6})(T - 20°) + (-0{\cdot}031 \times 10^{-6})(T - 20°)^2]$

In order to ascertain whether β is dependent on the past history of heat treatment, additional samples were heat-treated under production conditions in which longer times and slow heating and cooling rates were used. A sample was selected for test which gave an average temperature coefficient of almost zero. The resistance of this sample was then measured at 20° C intervals and the coefficients were determined. The results show a value for β of $-0{\cdot}031 \times 10^{-6}$ when $\alpha = -2 \times 10^{-6}$. As

Table 2

EFFECT OF HEAT TREATMENT OF 0·002 5 IN-DIAMETER EVANOHM ON THE TEMPERATURE COEFFICIENT OF RESISTANCE BETWEEN −60 AND +100° C

Heat-treating temperature	Time	Between −60 and 100° C			Between +20 and 100° C		
		Mean γ	α	β	Mean γ	α	β
565 C	0	+57	+57	−0·051	+54	+56	−0·025
	4 min	+35	+35	−0·045	+32	+34	−0·031
	7 min	+21	+21	−0·045	+18	+19	−0·023
	20 min	0	0	−0·031	−2	−2	−0·011
	1½ h	−16	−16	−0·025	−18	−17	−0·009
	27 h	+2	+2	−0·081	−5	−2	−0·044
	40 h	+17	+17	−0·123	+7	+13	−0·078
		Production heat-treated material (0·001 5 in)					
		−2	−2	−0·031	−4	−4	−0·012

All coefficients are expressed in parts in 10^6 per deg C. The mean heat-treating temperature is 565° C.

shown by the data in Table 2 for laboratory tests, $\beta = -0{\cdot}031 \times 10^{-6}$ over the temperature range −60–100° C when α is zero. Thus β does vary in a systematic manner during heat treatment, but is independent of past thermal history. The authors believe that the reason for the difference between these results and those reported previously lies in the greater uniformity obtained by heat treatment in a constant-temperature furnace instead of electrical heating.

(3) VARIABLES AFFECTING THE RESISTANCE OF HEAT-TREATED EVANOHM

After completion of heat treatment, the precautions necessary to evaluate appropriately the characteristics of resistance wire are many, depending on the type of resistor and the accuracy desired. The two biggest factors are the surface condition and the introduction of strain during fabrication of resistors.

(3.1) Surface Conditions

Stability of resistance at temperatures in the vicinity of 100–200° C is desirable. Quite frequently, tests conducted in air on bare wire at 100–200° C show excessive drift in resistance. When it is realized, however, that resistance is proportional to the square of the diameter of fine wire, even minute alterations, such as surface tarnishing, cause appreciable resistance changes. For example, if an oxide layer is formed on 0·004 8 in-diameter wire only $0{\cdot}16 \times 10^{-6}$ in thick, a change in resistance of 100 parts in 10^6 would be observed. Stability tests conducted in air at 190° C showed an initial high drift in resistance of about 0·1% after which the resistance was stable for 1 000 hours. By conducting a similar test in cracked ammonia, the initial drift was reduced to a few parts in 10^6. This difference was ascribed to surface oxidation.

It is believed that the initial drift in resistance of enamelled or coated wire is decreased because surface reactions are minimized.

(3.2) Effect of Strain on Electrical Properties of Evanohm

The effect of strain on Evanohm is responsible in many cases for the initial drift in resistance which is observed. In high-content nickel-chromium alloys, strain causes the resistivity to decrease. As shown by the curve in Fig. 2, this decrease is accompanied by an increase in temperature coefficient. The effect of strain on resistance is more difficult to evaluate. In simple tension, where uniform strain is obtained, the dimensional changes attending straining more than offset the decrease in resistance, and the change in resistance is always positive. In pure bending, the net dimensional changes are zero, and straining causes the change in resistance to be negative. In practice, resistors are constructed by applying a uniform tension (the winding tension) while the wire is coiled around a mandrel, bobbin or card. Thus both uniform and bending strains are introduced during fabrication. The net change in resistance then depends on which strains predominate, and the subsequent drift in resistance of a few parts in 10^6 reflect the relief or relaxation of stresses with time, during ageing at ambient or moderate temperatures. The authors are currently investigating in more detail the case of non-uniform strains produced during the winding of resistors. The general trend most frequently observed is in agreement with Arnold's observation that the temperature coefficient becomes more positive during winding. With due care during fabrication, these initial small spurious drifts in resistance can be minimized or essentially eliminated. Much additional information is necessary, however, to elucidate clearly the effects of non-uniform strain. Therefore, at present, it is appropriate to use extreme care in the fabrication of ultra-precision resistance devices in order to minimize these strains.

(4) REFERENCES

(1) Arnold, A. H. M.: 'Nickel-Chromium-Aluminium-Copper Resistance Wire', *Proceedings I.E.E.*, Paper No. 1084 M, July, 1956 (**103** B, p. 439).

(2) Nordheim, R., and Grant, N. S.: 'Resistivity Anomalies in the Nickel-Chromium System', *Journal of the Institute of Metals*, 1953–54, **82**, p. 440.

(3) Taylor, A., and Hinton, K. G.: 'A Study of Order-Disorder and Precipitation Phenomena in Nickel-Chromium Alloys', *ibid.*, 1952–53, **81**, p. 451.

621.317.794:536.3.08

Paper No. 2406 R
Sept. 1957

THE MEASUREMENT OF THERMAL AND SIMILAR RADIATIONS AT MILLIMETRE WAVELENGTHS

By G. R. NICOLL, B.Sc., Associate Member.

(*The paper was first received 3rd April, and in revised form 14th June*, 1957.)

SUMMARY

The measurement of thermal and similar noise radiations at millimetre wavelengths is discussed. It is shown how this type of measurement is applied to radiation from gas discharges, flames and crystal diodes, and how it is used in certain studies of the atmosphere. Two types of measuring instrument are compared.

(1) INTRODUCTION

There are certain similarities between the techniques used in the infra-red and microwave regions of the electromagnetic spectrum and these become more marked as radio techniques are developed at progressively shorter wavelengths. The basic source of energy used in making physical measurements in the visual and infra-red regions is the radiation emitted by any object at an elevated temperature. This thermal radiation is directly apprehended by ourselves in two ways, our eyes responding to the light from a fire and our skin to the heat. It is, however, not confined to these regions only, but occurs in some measure in all parts of the spectrum, and the possibility arises of using it as the source of energy in certain microwave measurements. In the radio-wave region the thermal radiation is quite minute compared with the oscillatory energy that may be generated by means of valve and circuit techniques; it therefore requires extremely sensitive radio-receivers to detect the radiation. In fact, thermal radiation at radio wavelengths is closely related in nature and magnitude to the random electrical fluctuations which occur in circuits and which are known as thermal or Johnson noise.

The relation between thermal radiation and Johnson noise can be illustrated by a simple model. A lossless transmission line which can propagate only one type of mode is terminated at the two ends by resistors R and R′, and for simplicity the resistances are chosen so that the line is matched at both ends. It is possible to show from general considerations, following the arguments of Nyquist,[1] that, if the two resistors are at the same temperature, T, a noise power kT per unit bandwidth is fed from one resistor into the line, along which it passes until it is absorbed by the other resistor. (k is the Boltzmann constant, $1{\cdot}38 \times 10^{-23}$ joule/deg K.)

An equal power is exchanged in the opposite direction, so that thermal equilibrium is maintained. If the temperatures of the resistors are not equal, but are T and T', then it is postulated that the temperature of one resistor does not affect the emission process of the other, with the result that noise powers kT and kT' per unit bandwidth are emitted respectively. This has been confirmed experimentally by Williams.[2] There is now a resultant exchange of energy from the hotter to the colder resistor, and the system tends to settle down at an equilibrium temperature intermediate between the two initial temperatures.

Now suppose that the resistor R is replaced by an aerial which is assumed to be matched to the line and that the aerial is surrounded by walls which are at a temperature T and which absorb completely any radiation falling upon them. As before, a noise power kT' per unit bandwidth is fed into the line by the resistor R′; the noise power passes along the line, to be radiated by the aerial and absorbed by the enclosing walls. Now the walls themselves emit thermal radiation into the enclosure, some of which will be intercepted by the aerial and will pass along the line to be absorbed by R′. In order to determine how much of the thermal radiation is absorbed by the resistor, the system is considered to be in thermal equilibrium, so that the resistor and the walls of the enclosure are at the same temperature T. In order that the resistor does not change its temperature (which would contravene the second law of thermodynamics), the flow of energy, in the form of thermal radiation, from the walls to the resistor must be exactly kT per unit bandwidth. If the temperature of the resistor is not equal to that of the walls, it is assumed that the resistor continues to absorb the same amount of energy, kT per unit bandwidth, from the walls, and to lose kT' per unit bandwidth to the walls. If the temperature of the walls is greater than that of the resistor, the latter will tend to become hotter, and if it is less the resistor will become colder.

No measurements made at R′ can distinguish whether the transfer of energy occurs between R′ and another resistor R, or between R′ and the walls of an enclosure. The model thus suggests that thermal radiation and Johnson noise are essentially the same. To put it another way, one normally thinks of Johnson noise as an electrical fluctuation occurring internally in some resistor, and of thermal radiation as a fluctuation in an external field; these are but two aspects of the same basic phenomenon.

The model with walls at a uniform temperature does not represent a case of much practical interest. Section 2 shows how to calculate the amount of thermal radiation absorbed by the aerial from an enclosure whose walls are not all at the same temperature.

It should be noted that the powers involved are extremely small; for example, if the walls of the enclosure are at room temperature—about 290° K—the power absorbed by the aerial at radio frequencies is 4×10^{-21} watt per unit bandwidth; and even if the receiver bandwidth is 10 Mc/s—which is a typical value for a wide-band superheterodyne receiver—the power amounts to no more than 4×10^{-14} watt.

Despite the minuteness of the powers involved, the detection and measurement of thermal radiation in the microwave region are of increasing importance. The interest originally lay in detecting the radiation from the sun, and the first successful results at microwave lengths appear to have been reported by Southworth.[3] He measured the radiation from the sun, at wavelengths in the range 1–10 cm, using conventional superheterodyne receivers of the type developed for radar systems. Dicke[4] introduced considerable improvements into the technique of detection, and the resulting instrument, called by him a 'radiometer', was able to detect, at wavelengths slightly greater than 1 cm, the radiation from the sun and the moon,[5] from the earth's atmosphere[6] and from nearby buildings.[7] Radiometers,

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Mr. Nicoll is at the Royal Radar Establishment.

using superheterodyne reception and incorporating Dicke's improvements, have been extensively used in the field of radio astronomy; in this application it is important that the instrument should be extremely stable over long periods. To this end, Machin, Ryle and Vonberg[8] have introduced further improvements in radiometers designed to work at about 100 Mc/s, but so far their improvements do not seem to have been applied to radiometers in the microwave band, no doubt because of technical difficulties.

There are two types of radiometer which have been used at millimetre wavelengths, the essential parts being shown schematically in Fig. 1. The first is the 'superheterodyne' radiometer, which is essentially similar to that used at longer wavelengths; rectification follows a change of signal to an intermediate frequency at which the signal is given pre-detector amplification. The second type is the 'straight' radiometer, which becomes practicable at millimetre wavelengths because r.f. bandwidths of thousands of megacycles per second can be achieved; in this type rectification is performed directly at the radio frequency and no intermediate-frequency amplification is used. The two types of radiometer are discussed in the Appendix, where it is shown that, while the 'straight' radiometer is a simple and relatively inexpensive instrument with sufficient sensitivity for many purposes, the superheterodyne radiometer is rather more sensitive and is more frequently used despite the additional expense and complication.

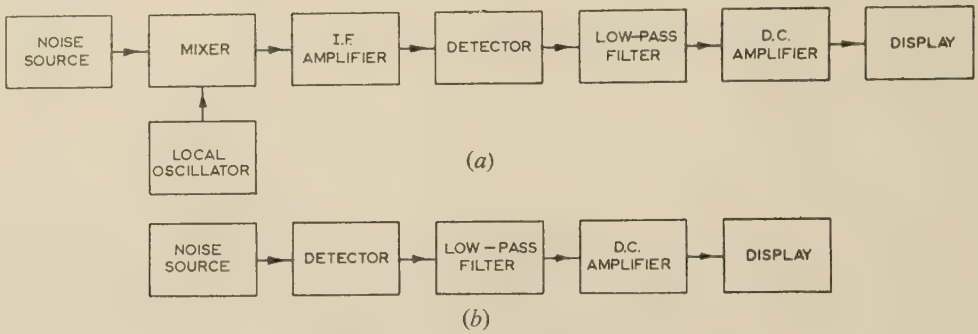


Fig. 1.—Basic radiometers.
(*a*) Superheterodyne.
(*b*) Straight.

An account is given in the paper of some of the measurements that have been made at millimetre wavelengths using radiometers of both types. The measurements described are chosen to illustrate the technique of millimetre radiometry and to show in general terms the kind of information that can be obtained from a study of thermal and other similar radiations.

One important measurement to be described is the calibration of the noise output of certain discharge lamps, which are used in determining the noise factors of radar receivers. In principle, the most satisfactory method of determining noise factor is the direct comparison of the noise produced internally in the receiver with the thermal noise in a resistor connected to the receiver input. The internal noise is large compared with the noise from the resistor if it is at room temperature, and to make the comparison simple it is desirable to heat the resistor to some thousands of degrees centigrade and so to make the thermal noise exceed the receiver noise. It is impracticable to heat a resistor to this extent, and there is used instead a discharge tube which emits the required amount of noise. It is necessary, however, to calibrate the noise output of the discharge tube, and this is done by comparing it with thermal noise by means of a radiometer; this comparison is described in Section 6. The possibilities of radiometry in various physical investigations are illustrated by the measurement of noise from flames and from crystal diodes (Sections 7 and 8).

The essential instrument used in radio astronomy is the radio telescope—a radiometer attached to an aerial. The calibration of the aerial for such an instrument is described in Section 9. Little work has yet been done in radio astronomy below a wavelength of one centimetre, where the possibilities of very fine angular resolution are difficult to apply because the atmosphere is only partially transparent. Clouds, water vapour, and even the oxygen of the air, absorb radiation from outer space and emit radiation characteristic of their own temperatures. If the highest precision is required in astronomical measurements at millimetre wavelengths, it will be necessary to build observatories on high mountains, as is already the practice with the larger optical telescopes.

These very troubles which interfere with astronomical investigations provide a method by means of which it is possible to study certain aspects of the structure of the atmosphere. This is illustrated by measurements of atmospheric attenuation described in Section 10.

Before the measurements are described it will be useful to summarize certain relations concerning the magnitude of the thermal radiation that is absorbed by an aerial and the limit in accuracy with which the radiation may be measured.

(2) COLLECTION OF THERMAL RADIATION BY AN AERIAL

It has been shown that an aerial surrounded by walls at a temperature T will absorb a power kT per unit bandwidth from the walls. A more practical situation is where the aerial receives energy from an object of finite extent and at a temperature T. Suppose the object is a black body, i.e. an object whose reflectivity is zero and whose emissivity is unity at all relevant wavelengths; let it subtend a solid angle $d\phi$ at the aerial. The magnitude of thermal or black-body radiation, and its variation with wavelength, are given by Planck's radiation law,[9] from which it can be shown that the energy flux falling normally through a surface of area A is given by

$$P = [2A\mathrm{F}(T)/\lambda^2]df d\phi \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\mathrm{F}(T) = hf/[\exp(hf/kT) - 1]$

f, λ = Frequency and wavelength of the radiation.

h, k = Planck and Boltzmann constants.

When this formula is applied to a microwave aerial such as a horn or parabolic mirror the effective surface area is not the physical area of the aperture, but is a related area known as the absorption cross-section, a. Only a portion of the energy

passing through the aperture of the aerial will be absorbed and the rest will be scattered; the scattering will be much more pronounced when the object does not lie on the axis of the aerial and the wavefront is no longer in phase across the aperture. The absorption cross-section will usually be less than the physical aperture, and for directions off the axis will be very much less. This is described in other terms by the well-known polar diagram of the aerial.

Thermal radiation is randomly polarized, but the energy can be considered to be divided, on the average, equally between two independent polarized components. The microwave receivers used in radiometers can respond to only one of these; therefore the energy absorbed by the receiver, assuming it to be matched, is half of that given by eqn. (1). For a general distribution of objects at different temperatures, the total power absorbed by the receiver is obtained by integrating the radiation coming from all directions, i.e.

$$P = \int_0^{4\pi} [a\mathrm{F}(T)/\lambda^2] df d\phi \quad . \quad . \quad . \quad . \quad (2)$$

A property of all aerials is that the total absorption cross-section, $\int_0^{4\pi} a d\phi$, is just λ^2. If the radiation falls uniformly on the aerial from all directions, corresponding to the enclosure at uniform temperature T, the total absorbed energy is

$$P = \mathrm{F}(T) df \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

For the microwave region of the spectrum, $hf \ll kT$ for nearly all practical values of the radiation temperature; under this condition kT may be used as an approximation for $\mathrm{F}(T)$. The error involved amounts to 5% when $f/T \simeq 2$ Gc/s per deg K; the error is thus becoming appreciable at the highest radio frequencies and the lowest temperatures, e.g. at 40 Gc/s and 20° K.

For the microwave region eqn. (3) may usually be replaced by

$$P = kTdf \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This is the result that was obtained by another argument in Section 1.

(3) A FUNDAMENTAL LIMIT IN THE ACCURACY OF A MEASUREMENT OF THERMAL RADIATION

Thermal radiation or noise consists essentially in random fluctuations in the flow of energy; the mean value of the flow has been discussed in the previous Section, but it is the fluctuation about the mean which limits the accuracy of a measurement of the mean. According to Gabor,[10] the r.m.s. fluctuation in the flow of thermal energy is equal to the mean. In his treatment he divides the frequency band and the time involved in a measurement into information cells of size, $\Delta f \Delta t = 1$, each cell containing a thermal energy kT. If thermal noise is averaged over a bandwidth f_b and a time t, it is equivalent to an average over $f_b t$ cells. Now if a measurement with an r.m.s. error of σ is repeated n times the average derived from the n measurements will have an r.m.s. error of $\sigma/\sqrt{n}$. Therefore the r.m.s. fluctuation in the measurement of thermal energy is $kT/\sqrt{(f_b t)}$, and if this is considered as being due to an apparent fluctuation ΔT in the temperature of the source,

$$\Delta T = T/\sqrt{(f_b t)} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

This, then, gives the limiting accuracy in the measurement of the temperature of the source. The limit can be made smaller by increasing the observation time, but since the variation is as $t^{-1/2}$, the rate of improvement is small; and, although it is possible to integrate over periods as long as one hour, it is not usually convenient to take much more than about a second or perhaps a minute for each measurement. The alternative is to increase the bandwidth of the noise being observed. Bandwidths of 10^7 c/s are commonly used in superheterodyne radiometers, and of nearly as much as 10^{10} c/s in 'straight' radiometers. With 1 sec integration, which is chosen as a standard for comparison, it is therefore fundamentally impossible to measure the temperature of a noise source to better than 1 part in 3000 and 1 part in 100000 respectively.

In practice it is found that unwanted noise generated within the radiometer prevents these limits being approached. The temperature resolution, ΔT, of the instrument is consequently independent of the temperature of the source, provided it is not too great; it is still proportional to $t^{-1/2}$.

Typical values for the temperature resolution at wavelengths just below 1 cm, and measured with an averaging time of 1 sec, are a few degrees centigrade for a superheterodyne radiometer and about 100° C for a straight radiometer.

(4) COMPARISON OF THERMAL RADIATION AT INFRA-RED AND MILLIMETRE WAVELENGTHS

To illustrate the relative magnitudes of thermal radiation at millimetre and infra-red wavelengths, consider the detection by receivers in these two wavebands of radiation from the same black-body source. Let the source subtend a small solid angle at the receivers and let it be at a temperature sufficiently great to make $\mathrm{F}(T) = kT$ at both wavelengths. The absorption cross-sections of the receivers are assumed to be equal, as also are the percentage bandwidths; $a\lambda df d\phi$ is therefore constant. The power absorbed, from eqn. (2), varies inversely with λ^3. Millimetre wavelengths are some 10^3 times greater than infra-red wavelengths, and consequently the infra-red radiation is 10^9 or so times greater than that at millimetre wavelengths. For wavelengths or temperatures which make $hf > kT$ the factor $\mathrm{F}(T)$ decreases exponentially as the frequency is increased; therefore, for temperatures near absolute zero the difference in the radiation powers at infra-red and millimetre wavelengths is not so marked.

Now compare the limiting sensitivity of millimetre radiometers with that of infra-red detectors: in both cases the limit is determined by a combination of noise from the detector and noise from the associated circuits. Infra-red photo-conductive cells, which are now widely used, have been reviewed by Smith.[11] It appears that the best results, using uncooled cells, are achieved with lead sulphide and the limiting sensitivity lies in the range 10^{-10} to 10^{-11} watt per c/s of integration bandwidth (for a cell area of 0·1 cm^2, which is taken as a standard).

A straight radiometer has an ultimate sensitivity of about 4×10^{-12} watt per c/s when current millimetre-wave diodes are used, and this is rather better than the value for the infra-red detector. A superheterodyne radiometer has a much lower ultimate sensitivity—a few times 10^{-16} watt per c/s—but this is not sufficient to overcome the tremendous difference in radiated energies between the two bands. Consequently, when they are compared in terms of temperature resolution, millimetre radiometers are much poorer than infra-red detectors. This difference in performance will be greatly reduced, and may possibly be reversed, when the radiating source is extremely cold—within a few degrees of absolute zero—because the exponential in the factor $\mathrm{F}(T)$ will reduce the radiation at infra-red by many orders.

(5) NOMENCLATURE

The term 'radiation temperature' or 'equivalent black-body temperature' is commonly used to describe the intensity of radiation. The concept is useful, and the following definitions have been used in the paper:

Radiation temperature.—The radiation temperature, measured at a given wavelength, is equal to the temperature of the black body which has the same angular extent as the source and which would produce the same intensity of radiation at the receiver.

Internal radiation temperature.—It is useful to distinguish an internal radiation temperature equal to the radiation temperature divided by the emissivity of the source. For a thermal radiator the internal radiation temperature is identical with the temperature of the radiating body; in some cases, e.g. crystal diodes, the internal radiation temperature will not be related directly to any thermal temperature.

Apparent radiation temperature.—When a radiating surface does not have an emissivity of unity, the radiation from other sources may be partially reflected by the surface, and then the total radiation, reflected and emitted, may be distinguished by an 'apparent radiation temperature'.

(6) RADIATION FROM A GAS DISCHARGE

It is now well known that an electrical discharge in a gas emits radiation in the microwave region of the spectrum. Provided that the type of gas, its pressure, the ambient temperature, the dimensions of the discharge and the current are specified, the magnitude of the radiation is reasonably constant and reproducible. It has been established, on both theoretical and experimental grounds, that the internal radiation temperature is closely related to the electron temperature in the discharge.[12, 13] The radiation from a discharge tube provides a convenient way of measuring the noise figure of radar receivers, but to do so it is necessary to calibrate the output from the discharge tube. This may be done, using radiometer techniques, by a direct comparison with the thermal radiation emitted by a resistive load placed in a waveguide and heated to some suitable temperature.*

The measurements to be described were made originally with a straight radiometer at a wavelength of about 8 mm, and it is believed that the consequent determination of absolute power at the micromicrowatt level was then more accurate than that at the milliwatt level.

The discharge tubes,[14] which were supplied by the Services Electronic Research Laboratory, were 0·1 in in diameter and were filled with neon at a pressure of 30 mm Hg. The tube was mounted slantwise across the waveguide so that the path length through the discharge for a wave in the guide was two or three wavelengths. The radiation from a discharge tube was about 50 times greater than from the heated waveguide load. Consequently, one tube was selected which, when run with 40 mA through the discharge, provided a high-level reference temperature, T_s, against which the other tubes could be compared; T_s was then determined carefully by comparison with the radiation from the waveguide load heated to about 200° C. A typical comparison is shown in Fig. 2. First the signal from the waveguide load was recorded for about 40 min using a time-constant in the radiometer of 16 sec; then the gain of the post-detector amplifier was decreased 50 times by means of a calibrated attenuator, and the signal from the discharge tube was recorded for a few minutes. In this way T_s was found to be 16 900° K, with an r.m.s. deviation of ±480° K on eight separate comparisons. Since the radiometer is linear to a high degree, the chief error arises from the noise in the radiometer when observing the waveguide load. This error could now be reduced if a later version of the waveguide load, which can be heated to 600° C, were used. The early forms of superheterodyne radiometer were not suitable for this type of measurement, because of spurious deflections on the recorder which could be produced by quite small changes in r.f. impedance of the source.

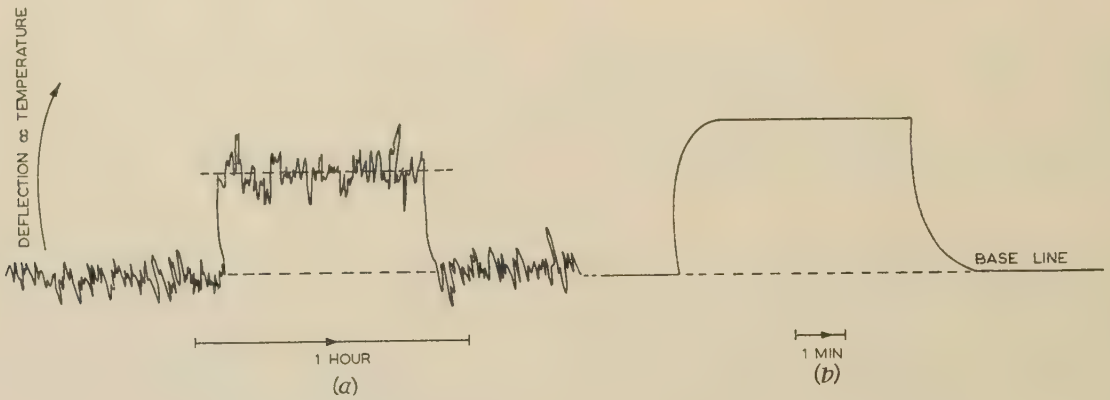


Fig. 2.—Calibration of radiation from a discharge lamp.
Base-line corresponds to radiation from matched load at room temperature.
(*a*) Signal from matched load 220° C above room temperature; amplifier at full gain.
(*b*) Signal from discharge lamp when lamp current was 40 mA; gain reduced 50 times and recorder speed increased.

These measurements give the radiation temperature for the discharge, but it is also of interest to determine the attenuation of the discharge and the internal radiation temperature; this can be done quite simply. Let the model shown in Fig. 3 represent the discharge. The attenuating region, whose power absorption coefficient is $a(z)$ and whose internal radiation temperature is $T_a(z)$, extends from $z = 0$ to $z = l$. Let radiation amounting to kT_b per unit bandwidth fall on the boundary at $z = 0$; it is desired to calculate the total radiation, kT_0 per unit bandwidth, emerging from the region at $z = l$. (Boltzmann's constant is now omitted and the calculation continues in terms of radiation temperature.) The radiation temperature comprises two components; T_0', which is the radiation T_b attenuated by the intervening layer, and T_0'', which is the thermal radiation emitted by the attenuating region. It will be assumed that no reflections occur at the boundaries of the region.

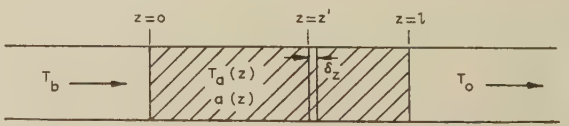


Fig. 3.—Radiation from an attenuating region.

Consider the effect of a small layer dz' at $z = z'$. The radiation

* Similar calibrations at wavelengths of 3 and 10 cm have been described.[25,26]

from the background is reduced to T' at $z = z'$ and suffers a further attenuation, $-dT'$ on passing through dz'.

Hence $$dT' = -a(z)T'dz'$$

Integrating this from $z = 0$ to $z = l$ gives

$$T_0' = T_b \exp\left(-\int_0^l adz\right)$$

The radiation produced by layer dz' is $(aT_a dz')$ and is attenuated in the region $z = z'$ to $z = l$, emerging as

$$dT'' = aT_a dz' \exp\left(-\int_{z'}^l adz\right)$$

The total thermal radiation from the layer is therefore

$$T_0'' = \int_0^l dT'' = \int_0^l aT_a \exp\left(-\int_{z'}^l adz\right)dz'$$

The total radiation emerging is

$$T_0 = T_0' + T_0'' = T_b \exp\left(-\int_0^l adz\right) + \int_0^l aT_a \exp\left(-\int_{z'}^l adz\right)dz' \quad (6)$$

If the internal radiation temperature and the absorption coefficient are constant throughout the region,

$$T_0 = T_b\varepsilon^{-al} + T_a(1 - \varepsilon^{-al}) \quad . \quad . \quad . \quad . \quad (7)$$

If the discharge in the tube is assumed to be uniform, it can be seen from the last equation that two measurements of T_0 for two separate values of T_b are sufficient to determine the internal radiation temperature T_a and the total attenuation ε^{-al}. This has been done by placing behind the discharge, first a matched waveguide load (giving $T_b \simeq 290°$ K), and secondly, a reflecting termination which effectively backs the lamp by its own image (giving $T_b = T_0$, where T_0 in this case is the value obtained with $T_b \simeq 290°$ K, since the radiometer is effectively a matched load at 290° K). The results of such measurements are shown in Fig. 4(*a*), where temperatures are quoted as fractions of the standard temperature, T_s, described above; the two solid lines show the measured radiation temperatures for the two kinds of background. The attenuations through the discharge have been calculated from these pairs of measurements and are found to agree, within experimental error, with a direct measurement of attenuation using a monochromatic source at about 1 mW level [Fig. 4(*b*)]. The internal radiation temperature of the discharge is then calculated, allowing for the small, zero-current attenuation which is due to loss in the waveguide and the glass envelope and which is assumed to be uniformly distributed along the discharge; the resultant temperature is shown by the broken line.

The radiation temperature of a discharge lamp was investigated over a range of wavelengths by inserting sections of waveguide which cut off all radiation greater than a certain wavelength determined by the width of the section. The cut-off wavelength was varied from 14 to 4 mm. The sensitivity of the crystal diodes deteriorated very rapidly with decreasing wavelength—more so, in fact, than could be caused by a constant barrier capacitance in the diode; this probably indicates that the crystal diode was badly mismatched at the shorter wavelengths. Even with a cut-off wavelength of 4 mm it was found possible to get an approximate calibration of the discharge lamp against radiation from a hot load; the radiation temperature was found

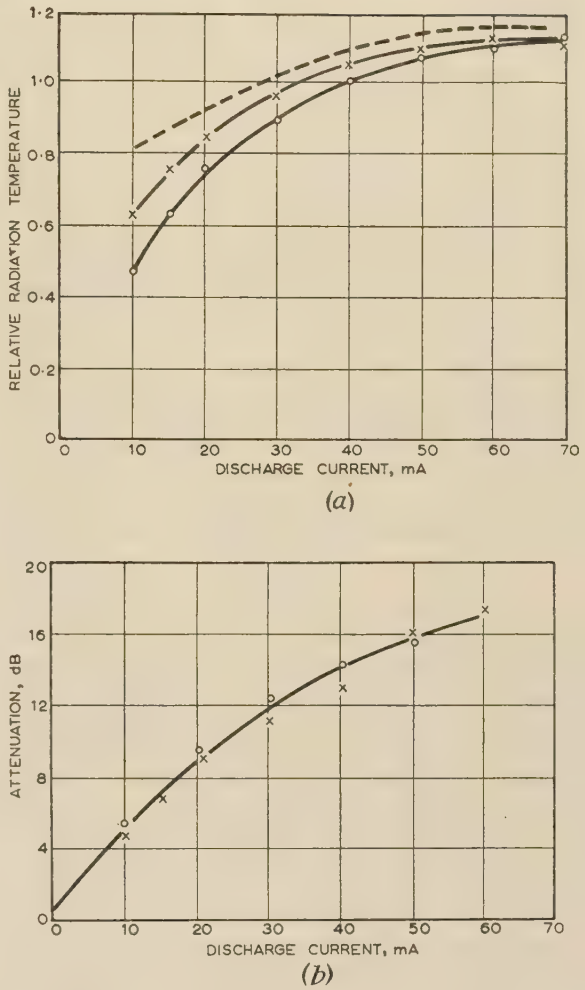


Fig. 4.—Variation of attenuation and radiation from current passing through a discharge lamp.

(*a*) Radiation temperature, in units of approximately 17 000° K.
○ Wooden load behind discharge.
× Reflecting piston behind discharge.
– – – Internal radiation temperature of the discharge, calculated from the radiation measurements.
(*b*) Attenuation in the discharge.
○ Measured directly.
× Calculated from the radiation measurements.

to be about four times less than the 8 mm value. This decrease was presumably due partly to a decrease in the attenuation of the discharge as the frequency was increased and partly to loss in the glass envelope and elsewhere.

(7) RADIATION FROM A FLAME

The kinetic energy of electrons in flames is known to be very large, corresponding to a high value for the temperature of the electron gas. Consequently it is to be expected that a flame will emit microwave radiation in a manner similar to an electrical discharge tube.

An oxygen and coal-gas flame was examined for radiation in a band centred at about 8mm. The radiation from the flame, produced by means of a large burner, was collected by a lens of 6 in diameter and 8 in focal length. In the absence of the flame the deflection of the recorder corresponded to the background temperature produced by the walls of the room; when the flame was lit no change in the deflection could be observed, but when the flame was primed with potassium and sodium carbonates a signal of about 200° C was observed. The inter-

pretation is that the total attenuation through the normal flame was extremely low and the contribution to the radiation from the flame was negligible; when the flame was primed, the electron density was increased through ionization of the salts, and this produced an appreciable amount of attenuation, with an associated increase in radiation. The signal was not steady, because of the method of priming, and no attempt was made to determine the internal radiation temperature of the flame.

The technique seems to offer a promising method of studying one aspect of the physics of flames and could be improved considerably. For example, the flame could be placed between two horns or lenses, one transmitting radiation from a discharge lamp and the other receiving the radiation from the lamp through the flame. From the signals received with the lamp on and with the lamp off, the attenuation and the internal radiation temperature of the flame could be calculated. Alternatively, an extinction technique could be employed by adjusting the radiation from the lamp by means of an attenuator until the same signal was received with and without the flame; the background temperature would then be equal to the internal radiation temperature of the flame.

(8) RADIATION FROM A CRYSTAL DIODE

Microwave crystal diodes emit noise or radiation when a bias current is passed through them. Some measurements have been made on millimetre-wave diodes;[15] no radiation was observable for bias currents in the low-impedance direction, but in the reverse direction the radiation temperature was found to increase nearly linearly with bias current. Temperatures of up to 2000° K have been observed for bias currents of 1mA. The results are consistent with the radiation being shot noise produced by the current carriers crossing the potential barrier which is assumed to exist between the metal whisker and the semi-conductor material.

(9) CALIBRATION OF AN AERIAL

In connection with the atmospheric and astronomical observations, which will be described in Section 10, it was necessary to determine the properties of the aerial used. When measurements of the gain and the polar diagram are being made on narrow-beam high-gain aerials it is essential to make sure that any auxiliary aerial used is sufficiently far away to be in the Fraunhofer region of the principal aerial, which can then be considered to be a point source. The region near the aerial is the Fresnel region, where simple optical theory may be used. If the aerial consists of a parabolic mirror with a small horn-feed at its focus, the energy is radiated in the Fresnel region as a cylindrical beam having the same radius, R, as the parabolic mirror; further out, diffraction predominates as the Fraunhofer region is entered, and the energy is radiated chiefly within a cone of semi-angle $\lambda/2R$. A convenient measure of the transition from one region to the other is that distance, z, at which the surfaces of the cylinder and the cone intersect. This gives

$$z\lambda/R^2 = 2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

z may be considerable for large aerials at millimetre wavelengths.

In order to test this simple picture, a source of radiation was placed on the axis of the aerial at varying distances from the aerial. When the source was in the Fraunhofer region, the signal received varied inversely with the square of the distance; when it was near the aerial, in the Fresnel region, the signal received was independent of the distance. The transition region corresponded roughly to distances between $z\lambda/R^2 = 1$ and $z\lambda/R^2 = 2$.

To determine the gain of the aerial the following method was employed. Two aerials were used, one as a transmitter and the other as a receiver, with gains g_T and g_R respectively. If P_T is the power radiated by the first aerial, P_R is the power absorbed by the second and the separation is z, then, provided that the separation is great enough for the inverse square law to hold,

$$P_R = P_T g_T g_R (\lambda/4\pi z)^2 \quad . \quad . \quad . \quad . \quad (9)$$

By measuring P_R/P_T, knowing λ and z, for three aerials taken in pairs, it is possible to determine the gain of each aerial.

The three aerials used were the paraboloid mirror, a short circular horn and a long rectangular horn. A discharge tube was used as the transmitter and a calibrated attenuator was used for determining power ratios. The stability of the system was very good. The separation of each pair of aerials was sufficient to ensure that it was greater than the transition distance for the aerial with the larger gain.

For the paraboloid mirror, which was 44 in in diameter with a focal length of 18·7 in, the gain was found to be 100000 with an estimated error of ±5%; this is in good agreement with a calculated value.

In addition, the shape of the main lobe in the polar diagram of the paraboloid aerial was measured.

(10) ATMOSPHERIC ATTENUATION STUDIES

Microwave radiometry offers a powerful and simple means of studying the attenuation of microwave radiation in the atmosphere, two general methods being employed. The first makes use of the radiation from a celestial body, such as the sun. If, for example, the radio telescope is pointed at the sun, it is found that the radiation reaching the telescope is diminished when a cloud passes between the telescope and the sun. It is therefore possible to determine the change in the total attenuation produced by various types of cloud.* If the sun is observed during the course of the day at different zenith angles, it is found that the radiation received is smallest when the sun is near the horizon and largest when it is near the zenith, corresponding to the attenuation produced by the varying path-length through the atmosphere. If the atmosphere is assumed to be horizontally stratified, it is possible to determine from the observations the total attenuation at any specific zenith angle, by means of a simple relation between zenith angle and path length. It is necessary to assume that neither the radiation from the sun nor the composition of the atmosphere changes during the sequence of observations.

The second method, which was first employed at infra-red wavelengths and was introduced at microwavelengths by Dicke *et al.*,[6] makes use of the thermal radiation from the atmosphere itself. The amount of this radiation depends on the attenuation of the atmosphere, in a manner similar to that described for a gas discharge in Section 6. The path length through the atmosphere, which is again assumed to be horizontally stratified, is proportional to the secant of the zenith angle of the path, and the total vertical attenuation may be deduced from the variation of the radiation with zenith angle. This method is more satisfactory than the first, because it is possible to complete the necessary measurements in a matter of minutes, whereas in the first it takes more than an hour for the zenith angle of the celestial body to vary by an amount sufficient for accuracy. During the longer period the method may be invalidated by changes in the composition of the atmosphere.

The two celestial bodies of interest in this connection are the sun[16] and the moon.[17] According to measurements made with the present radiometers, using aerials whose gains were determined experimentally and theoretically, the sun has a radiation temperature of about 10000° K at a wavelength of 8mm; the radiation is sufficiently constant to be used for attenuation

* Contributions to the signal from frequencies higher than the fundamental band to which the radiometer is tuned and from direct heating effects may be neglected.

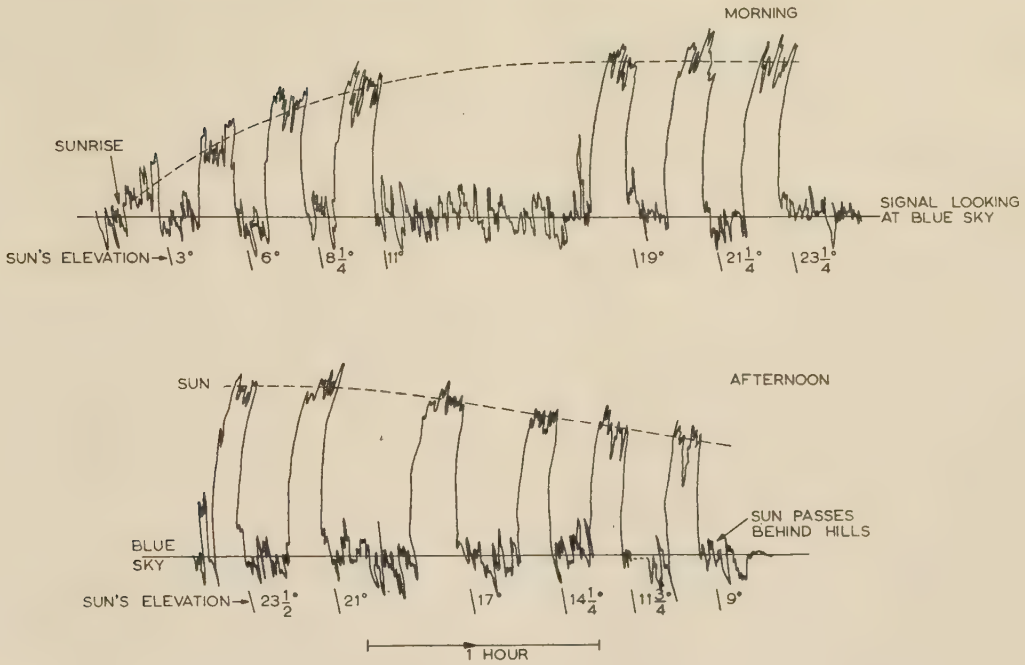


Fig. 5.—Radiation from the sun, from sunrise to sunset, 12th October, 1950.
Vertical deflection is proportional to radiation temperature.

measurements. The radiation temperature from the moon is, on the average, about a fortieth of that from the sun and shows a sinusoidal variation during the lunar month with a peak-to-peak amplitude of about one-sixth of its mean value. The maximum temperature occurs exactly an eighth of a lunar month after full moon, as was discovered at a wavelength of 12·5 mm by Piddington and Minnett;[17] this 'phase lag' of 45° has been verified at a wavelength of 8 mm. The moon provides a source of radiation which could be used for attenuation measurements at night. However, the radiation from the moon is so small that a radiometer sufficiently sensitive to measure it would also be capable of detecting the radiation from the atmosphere and the better method proposed by Dicke could then be employed. The sun's radiation is sufficiently great to enable measurements to be made when a radiometer capable of detecting the atmospheric radiation is not available.

Various studies of atmospheric attenuation have been made by both these methods at 8 mm, and preliminary observations, using the sun, have been made at 4 mm.

The techniques are illustrated in Figs. 5–7. Fig. 5 shows the changing radiation from the sun during a day in which the atmosphere remained relatively stable. A radio telescope, comprising a paraboloid mirror 44 in in diameter and a 'straight' radiometer centred on 8 mm wavelength, was pointed alternately at the sun and at the blue sky beside it, giving the recording shown. The maximum aerial temperature due to the sun was about 3 000° K; the fluctuations were the inherent noise in the radiometer, together with a certain amount of microphony. The signal from the sun was completely attenuated when the sun was close to the eastern horizon and increased rapidly during the two hours following sunrise. During most of the day the signal was fairly steady, but it began to fall rapidly about two hours before sunset. The sun could not be followed to the western horizon, because it passed behind some hills at an elevation of 10°.

A typical run observing atmospheric radiation is shown in Fig. 6. A superheterodyne radiometer was used; zero deflection

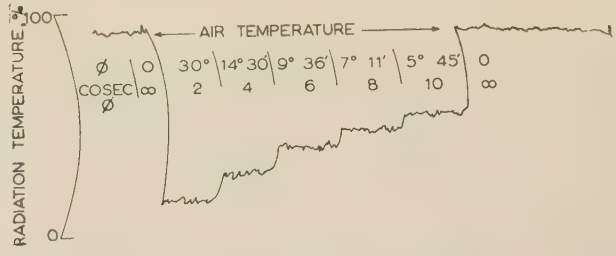


Fig. 6.—Radiation from a clear atmosphere, showing a total vertical attenuation of 0·36 dB.

in the recording corresponds approximately to 0° K and full deflection to about 320° K. The radio telescope was first pointed at the horizon ($\phi = 0°$, where ϕ is the complement of the zenith angle) and a deflection corresponding to the air temperature was obtained, since the path length through the atmosphere, and hence the attenuation, was very great. The telescope was then pointed successively at elevations which gave path lengths of 2, 4, 6, 8 and 10 times the vertical path length through the atmosphere. The sky was blue everywhere, and the total vertical attenuation of 0·36 dB for the clear atmosphere was deduced from the measurements.

A large number of measurements of the attenuation produced by clear atmospheres at a wavelength of 8·2 mm were made over a period of nearly a year. The total vertical attenuation was found to range between $\frac{1}{4}$ and $\frac{1}{2}$ dB, depending on the amount of uncondensed water vapour in the atmosphere. The lower value is applicable to a winter's day when it is freezing at ground level, and the higher value to a summer's day with the ground temperature in the region of 80° F. The humidity of the air during each measurement was determined from the standard radio-sonde data published daily by the Meteorological Office. The results indicated an absorption due to oxygen of about 0·21 dB, and the variation with humidity was in fair agreement with the results of Becker and Autler.[18]

When light cloud was present the attenuation was found to increase relative to the clear-sky attenuation by up to $\frac{1}{10}$ dB. Heavy cloud was found to produce attenuation of up to nearly 3 dB, as illustrated by Fig. 7, which shows the absorption of the

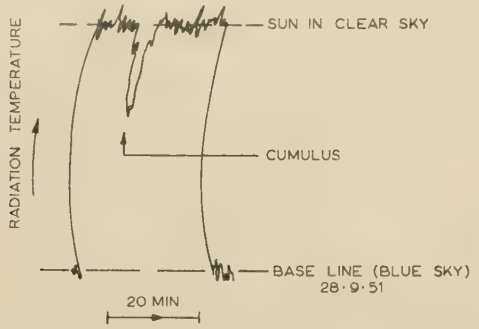


Fig. 7.—Attenuation of the sun's radiation by an isolated cloud.
Cumulus with strong vertical development, base at about 3 000 ft, top about 8 000 ft.

radiation from the sun when a cumulus cloud with a strong vertical development passed between the sun and the radiometer. The recording was made with a straight radiometer. This attenuation gives a possible method of estimating the total amount of condensed water contained in a vertical column within the cloud.

Rain was found to produce vertical attenuations ranging from $\frac{1}{2}$ to 1 dB for light rain up to 10 dB for very heavy rain.

(11) CONCLUSIONS

Radiometers for millimetre wavelengths have now been developed to the stage where they are robust and stable instruments capable of making measurements with considerable precision. They have been used for examining radiation from crystal diodes, electrical discharges, the atmosphere and celestial bodies.

The techniques used in millimetre radiometry show a certain resemblance to methods used at optical and infra-red wavelengths. The gap between infra-red and radio wavelengths is rapidly being closed; one instance of this is the present observations of the sun at 4 mm with a radio receiver and the observations made in recent years by Sinton[19, 20] at wavelengths between 1 and 3 mm using a Golay cell.*

(12) ACKNOWLEDGMENTS

The work which has been described could not have been carried out without the many contributions from the author's colleagues, especially on the design and construction of valves and other millimetre components. In particular the author wishes to thank Mr. F. Wood, who has assisted in various measurements, and especially Mr. F. L. Warner, who has contributed throughout to all aspects of the work.

The paper is published by permission of the Controller, H.M. Stationery Office.

(13) REFERENCES

(1) Nyquist, H.: e.g. see Goldman, S.: 'Frequency Analysis, Modulation and Noise' (McGraw-Hill, New York, 1948), p. 385.
(2) Williams, F. C.: 'Coexistent Thermal and Thermionic Fluctuations in Complex Networks', *Journal I.E.E.*, 1938, **83**, p. 76.
(3) Southworth, G. C.: 'Microwave Radiation from the Sun', *Journal of the Franklin Institute*, 1945, **239**, p. 285.
(4) Dicke, R. H.: 'The Measurement of Thermal Radiation at Microwave Frequencies', *Review of Scientific Instruments*, 1946, **17**, p. 268.
(5) Dicke, R. H., and Beringer, R.: 'Microwave Radiation from the Sun and Moon', *Astrophysical Journal*, 1946, **103**, p. 375.
(6) Dicke, R. H., Beringer, R., Kyhl, R. L., and Vane, A. B.: 'Atmosphere Absorption Measurements with a Microwave Radiometer', *Physical Review*, 1946, **70**, p. 340.
(7) Lawson, J. L., and Uhlenbeck, G. E.: 'Threshold Signals', Volume 24 in Radiation Laboratory Series (McGraw-Hill, New York, 1950), p. 107.
(8) Machin, K. E., Ryle, M., and Vonberg, D. D.: 'The Design of an Equipment for Measuring Small Radio-frequency Noise Powers', *Proceedings I.E.E.*, Paper No. 1263 R, May, 1952 (**99**, Part III, p. 127).
(9) Roberts, J. K.: 'Heat and Thermodynamics' (Blackie, London, 1940), Chapters XIX and XX.
(10) Gabor, D.: 'Communication Theory and Physics', *Philosophical Magazine*, 1950, **41**, p. 1161. See also some corrections given in 'Lectures in Communication Theory', Massachusetts Institute of Technology, 1952, Technical Report No. 238.
(11) Smith, R. A.: 'Infra-Red Photo-Conductors', *Philosophical Magazine Supplement*, 1953, **2**, p. 321.
(12) Mumford, W. W.: 'A Broad-Band Microwave Noise Source', *Bell System Technical Journal*, 1949, **28**, p. 608.
(13) Knol, K. S.: 'Determination of the Electron Temperature in Gas Discharges by Noise Measurements', *Philips Research Reports*, 1951, **6**, p. 288.
(14) Bridges, T. J.: 'A Gas-Discharge Noise Source for Eight-millimetre Waves', *Proceedings of the Institute of Radio Engineers*, 1954, **42**, p. 818.
(15) Nicoll, G. R.: 'Noise in Silicon Microwave Diodes', *Proceedings I.E.E.*, Paper No. 1671 R, September, 1954 (**101**, Part III, p. 317).
(16) Pawsey, J. L.: 'Solar Radio-Frequency Radiation', *ibid.*, Paper No. 839 R, September, 1950 (**97**, Part III, p. 290.)
(17) Piddington, J. H., and Minnett, H. C.: 'Microwave Thermal Radiation from the Moon', *Australian Journal of Scientific Research*, 1949, *A*, **2**, p. 63.
(18) Becker, G. E., and Autler, S. H.: 'Water Vapour Absorption of Electromagnetic Radiation in the Centimeter Wavelength Range', *Physical Review*, 1946, **70**, p. 300.
(19) Sinton, W. M.: 'Detection of Millimetre Wave Solar Radiation', *ibid.*, 1952, **86**, p. 424.
(20) Sinton, W. M.: 'Observations of Solar and Lunar Radiation at 1·5 Millimetres', *Journal of the Optical Society of America*, 1955, **45**, p. 975.
(21) Torrey, H. C., and Whitmer, C. A.: 'Crystal Rectifiers', Radiation Laboratory Series, Vol. 15 (McGraw-Hill, New York, 1948), Chapter 11.
(22) Goldman, S.: 'Frequency Analysis, Modulation and Noise' (McGraw-Hill, New York, 1948), p. 242.
(23) Ditchfield, C. R.: 'Crystal Mixer Design for Frequencies from 20 000 to 60 000 Mc/s', *Proceedings I.E.E.*, Paper No. 1548 R, November, 1953 (**100**, Part III, p. 365).
(24) Mayer, C. H.: 'Improved Noise Power Measurements through the Use of Ferrites', Report of the Washington Conference on Geophysical Research, 1954.
(25) Hughes, V. A.: 'Absolute Calibration of a Standard Temperature Noise Source for Use with S-band Radiometers', *Proceedings I.E.E.*, Paper No. 2125 R, September, 1956 (**103** B, p. 669).
(26) Sutcliffe, H.: 'Noise Measurements in the 3-cm Waveband using a Hot Source', *ibid.*, Paper No. 2130 R, September, 1956 (**103** B, p. 673).

(14) APPENDICES

The 'Straight' Radiometer

In the 'straight' radiometer, the radiation is rectified by means of a silicon-tungsten point-contact diode, which for signals below a level of about 1 μW acts as a square-law detector.[21] Thus

$$I = \beta P \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where I is the rectified output current, P is the input power and β is the rectification efficiency of the diode. The mean-square fluctuation of the output current is given by

$$\overline{I^2} = N_d 4kTf_i/R \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where N_d is the noise ratio of the diode, T is its temperature, R is its slope resistance and f_i is the bandwidth of the integrating circuit.

* A Golay cell detects radiation by the expansion produced in a gas due to the heating effect of the absorbed radiation.

The minimum detectable signal, P_{min}, is taken conventionally as being equal to the r.m.s. fluctuation, although a more realistic value is 6–8 times greater.[22] This gives

$$P_{min} = (N_d 4kTf_i/R)^{1/2}/\beta \quad . \quad . \quad . \quad (12)$$

A convenient figure of merit for comparing various diodes is

$$M = \beta(R/N_d)^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Thus

$$P_{min} = (4kTf_i)^{1/2}/M \quad . \quad . \quad . \quad . \quad . \quad (14)$$

$$= 1{\cdot}3 \times 10^{-10} f_i^{1/2}/M \text{ watt} \quad . \quad . \quad (15)$$

If radiation from a load at T_L is detected by means of the diode, a change, ΔT_L, in temperature of the load will produce a change $k\Delta T_L f_n$ in the power absorbed by the diode, where f_n is the r.f. bandwidth over which the diode is matched to the load. The smallest change in temperature of the load that can be observed is obtained by equating the above power to P_{min}, giving

$$\Delta T_L \simeq 10^{13} f_i^{1/2}/Mf_n \quad . \quad . \quad . \quad . \quad (16)$$

The figure of merit, M, is a very variable parameter. Values of up to 30 were found at a wavelength of 8 mm, but at longer wavelengths much larger values are possible. The sensitivity of the radiometer is very dependent on the r.f. bandwidth, but greater than 4 Gc/s at a central frequency of 40 Gc/s has been obtained from the waveguide-type diode described by Ditchfield.[23]

With $f_n = 4$ Gc/s and $M = 30$ the temperature sensitivity is

$$\Delta T_L = 80\sqrt{f_i} \text{ deg C} \quad . \quad . \quad . \quad . \quad . \quad (17)$$

The noise level from the diode, when no bias current is passed through it, corresponds to Johnson noise[15] (i.e. $N_d = 1$), and for a diode whose resistance is 10 kilohms this amounts to about 10^{-9} volt r.m.s. for a bandwidth of 1 c/s. The amplifier must be capable of working down to this level.

The radiation is interrupted in the usual manner by means of a resistive disc rotating in and out of a slotted waveguide. This enables an a.c. amplifier to be used; in practice, this was centred on 20 c/s. By selecting the first valve in the amplifier for low noise level and running under optimum conditions, it was found possible to make the equivalent noise resistance, at the grid, less than 20 kilohms; diode impedances were smaller than this, usually lying between 1 and 10 kilohms. A tuned transformer with a secondary impedance of 5 megohms, giving a voltage step-up of either 10 : 1 or 30 : 1, was used between the diode and this valve to ensure that the noise from the first valve was unimportant. The rest of the amplifier was conventional, consisting of *RC*-coupled stages followed by a phase-sensitive detector and an integrating circuit. Great care had to be taken to eliminate all sources of interference, including microphony and stray earth-currents. In order to define the band within which measurements were made, a suitable wide-band filter consisting of a half-wavelength section of low-impedance waveguide was used.

The sensitivity of the 'straight' radiometer is determined by the figure of merit of the diode. The diodes were assembled with as light a pressure on the whisker as was consistent with stability of the contact. Passing bias current through the diode reduced the figure of merit drastically. The effect of cooling the diode was examined, and it was concluded that this was not a promising method of improving the performance. An improvement could be obtained by r.f. amplification of the radiation. The noise level of the r.f. amplifier is not important in this type of radiometer so long as the sensitivity is limited by noise from the diode. Suitable amplifiers for millimetre wavelengths were not available.

(14.2) The Superheterodyne Radiometer

The techniques used in the superheterodyne radiometer need not be described in detail, for the theory has been given by Dicke.[4] The main trouble to be overcome arises from spurious signals caused by changes in the r.f. impedance presented to the mixer. In one form of radiometer a reflection coefficient of 0·1 at the aerial was found to produce a spurious change in the zero level of ±40°C as the phase of the reflection was varied, the error being proportional to the reflection coefficient. This trouble was eliminated by locking the frequency of the klystron local oscillator to that of a reference cavity. Mayer[24] has described how to overcome this effect by means of a ferrite isolator.

A certain amount of work was done on radiometers for wavelengths less than 8 mm, but the techniques were not sufficiently well established for reliable and repeatable measurements. Local oscillators at the required frequencies were not available and some form of harmonic operation was necessary. When the mixer was driven by 8 mm power the best noise factor achieved at 4 mm with reasonable stability was 24 dB. With separate distorter and mixer diodes the noise factor was about 20 dB. These values may be compared with 17 dB at 6·6 mm and 12 dB at 8 mm for the direct operation of balanced mixers driven at the fundamental frequency by klystron oscillators.

(14.3) Comparison of the Two Types of Radiometer

A 'straight' radiometer was more simple and less expensive than a superheterodyne one; it required few waveguide components, no klystron and only simple audio-frequency circuits. A 'straight' radiometer had sufficient temperature resolution for measurements of noise temperature of a few hundred degrees or more, but a superheterodyne radiometer was essential for measurements to an accuracy within a few degrees. The frequency resolution of a 'straight' radiometer was poor, about 1 part in 10, whereas the superheterodyne radiometer could do better than 1 part in a 1 000. The linearity of the 'straight' radiometer was excellent, even for very large signals, whereas the superheterodyne was linear only for signals much smaller than the receiver noise. As the wavelength was decreased the performance of both types deteriorated drastically, and even at 4 mm the temperature resolution of the 'straight' radiometer was found to be inferior to that of its superheterodyne counterpart.

DISCUSSION ON

'DESIGN OF MICROWAVE FILTERS WITH QUARTER-WAVE COUPLINGS'*

Mr. J. Horna (*Czechoslovakia: communicated*): The paper describes a very useful kind of inductive susceptance. We often need to construct a filter at some different frequency, i.e. for various values of $2a/\lambda_g$. Then it is convenient to plot the behaviour of the susceptance in normalized parameters d/a, $2a/\lambda_g$, $B(2a/\lambda_g)$. In accordance with the equation

$$-B\frac{2a}{\lambda_g} = \frac{8}{\log_\varepsilon\left(\frac{1}{12\cdot33}\frac{a}{d}\right) + 0\cdot0404\left(\frac{a}{\lambda}\right)^2}$$

the susceptance parameter $B(2a/\lambda_g)$ seems nearly independent of frequency and appears linearly on semilog paper (Fig. A).

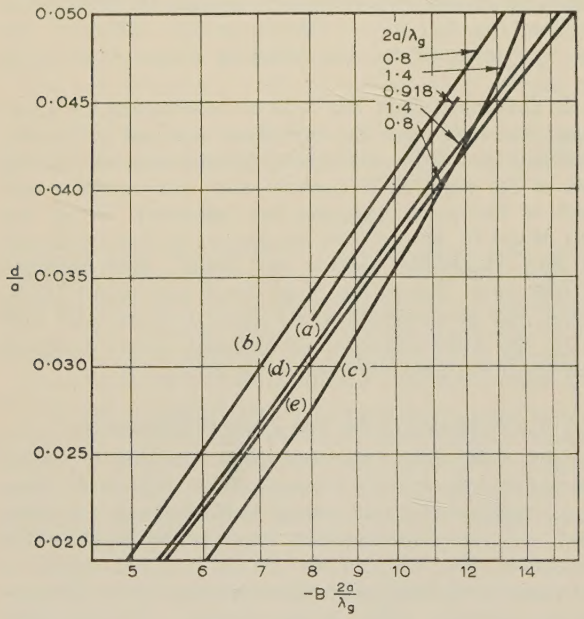


Fig. A

(*a*) Craven and Lewin's measurement.
(*b*), (*c*) Horna's measurements.
(*d*), (*e*) Theoretical values.

But our measurements at various frequencies show that we cannot rely upon low frequency-dependence even in the 3 800–4 200 Mc/s band; for $d/a = 0\cdot04$ the error would be about 10%, and it would increase to −30% at the highest recommended frequency of the rectangular waveguide. In accordance with the results the coefficient at $(a/\lambda)^2$ should be negative. We cannot improve the value of the coefficient from the measured points as its value would be dependent on d/a. The reduction of frequency dependence at $d/a = 0\cdot05$ may prove the influence of the longitudinal elements of the equivalent T-circuit. The theoretical results should be compared with the value of the transverse element of the equivalent circuit. For this reason it would be necessary to use the indirect measuring method by means of a movable plunger.[E]

At $d/a < 0\cdot03$ the measured cavity has too large a bandwidth and the frequency dependence of the susceptance parameter inside the pass band cannot be neglected. Therefore we have made these cavities longer by several half-wavelengths; in the numerical work we can easily include the increased selectivity of the line. Taking the desired properties of the filters into consideration we may say that a single post is sufficient when the value of b is small.[F]

The behaviour of the filter above the pass band is a valuable advantage in using the triplets of inductive posts. The quarter-wave coupling lines being short, the unwanted transparencies[G] are not closer to the pass band than with direct-coupled filters. At higher frequencies the triplets can act as mode filters.

References

(E) Storer, J. E., Sheingold, L. S., and Stein, S.: 'A Simple Graphical Analysis of a Two-Port Waveguide Junction', *Proceedings of the Institute of Radio Engineers*, 1953, **41**, p. 1004; correspondence, *ibid.*, 1954, **42**, p. 1447.

(F) Levy, R.: Previous discussion on Craven and Lewin's paper, *Proceedings I.E.E.*, 1956, **103** B, p. 632.

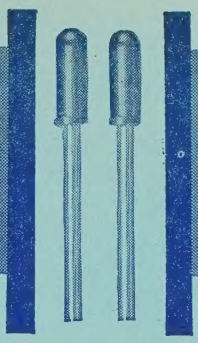
(G) Fedida, S.: 'Wide Band Microwave Radio Links', *Marconi Review*, 1955, **18**, p. 69.

Messrs. G. Craven and **L. Lewin** (*in reply*): We are very interested to see the comparison of our theoretical results with Mr. Horna's measurements. Our experience has been that the agreement should be good for small post diameters, the discrepancy becoming greater the thicker the posts. In our opinion the cause of the discrepancy must be sought, in part, in the inadequacy of the analysis for other than very thin posts. It is obvious from the formula for B that an infinite value would be reached for d somewhat greater than $a/12\cdot33$. This, of course, is not possible, so that the formula must break down before such a value is reached. Now, for a single post such a diameter is not excessive and the simple first-order theory gives a reasonable estimate of the susceptance. We conclude, therefore, that for a triplet of posts the second-order theory must be applied earlier at an appreciably smaller post size, since already at $d/a = 0\cdot08$ the formula is useless. The second-order theory requires an equivalent T-network so that the representation by a simple shunt susceptance itself is inadequate. We have not tried to extend our theory to include second-order effects: we believe that such an extension is straightforward in principle but is likely to lead to much heavy algebra in the analysis.

The value of triplets acting as mode filters for higher frequencies is a point which we had not considered but which is, perhaps, worthy of note.

* Craven, G., and Lewin, L.: Paper No. 2001 R, March, 1956 (see 103 B, p. 173).

TRANSISTOR NEWS



from Mullard

OC70 and OC71 new high performance

Maximum working collector voltage tripled . . . . . maximum junction temperature raised to 75°C . . . . . maximum power output correspondingly increased. These features characterise the increased performance of the new versions of the OC70 and OC71 junction transistors—increases which are made possible by further improvements in transistor production techniques backed by Mullard experience in the manufacture of many hundreds of thousands of transistors.

The new OC70 and OC71 are offered for use at collector voltages of 30V peak and 20V d.c. in both grounded base and grounded emitter where the circuit base resistance to earth is less than 500 ohms.

The increased maximum junction temperature now allowed may be exploited by circuit designers either in operating the OC70 and OC71 at the original power dissipations at 20°C higher ambient temperatures or in increasing the power dissipation at the original temperature. Some low power circuits can now be developed to work at the high *ambient* temperature of 70°C.

A revised data sheet including a curve showing the maximum d.c. collector voltage against external base-emitter circuit resistance is available from the address below.

ABRIDGED DATA *

Max. junction Temperature	75°C
Junction temp. rise above ambient in free air	0.4°C/mW
Max. collector dissipation at 45°C	75mW
Max collector dissipation at 25°C	125mW
Max. collector voltage, grounded base or grounded emitter:	
$V_{c\,(pk)}$ max.	–30V
V_c max.	–20V

* *Common characteristics for OC70 and OC71 general purpose and audio pre-amplifier transistors.*

MULLARD LIMITED, COMMUNICATIONS AND INDUSTRIAL VALVE DEPARTMENT

Mullard

MULLARD HOUSE
TORRINGTON PLACE, LONDON, W.C.1

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING) SEPTEMBER 1957

CONTENTS

Declaration on Fair Copying.—Within the terms of the Royal Society's Declaration on Fair Copying, to which The Institution subscribes, material may be copied from issues of the *Proceedings* (prior to 1949, the *Journal*) *which are out of print and from which reprints are not available.* The terms of the Declaration and particulars of a Photoprint Service afforded by the Science Museum Library, London, are published in the *Journal* from time to time.

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and year of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

THE BENEVOLENT FUND

The Object of the BENEVOLENT FUND is to help those members of The Institution and their dependants who have suffered a setback through ill-health, or who are faced with difficult circumstances.

HOW TO HELP THE FUND

Annual Subscriptions, preferably under deed of Covenant

Donations

Legacies

Subscriptions and Donations may be sent by post to

THE HONORARY SECRETARY
THE INCORPORATED BENEVOLENT FUND OF THE INSTITUTION OF ELECTRICAL ENGINEERS, SAVOY PLACE, W.C.2

or may be handed to one of the Local Honorary Treasurers of the Fund.

Though your gift be small, please do not hesitate to send it

LOCAL HON. TREASURERS OF THE FUND:

EAST MIDLAND CENTRE . . . . . *R. C. Woods*
IRISH BRANCH . . . . . *A. Harkin, M.E.*
MERSEY AND NORTH WALES CENTRE *D. A. Picken*
NORTH-EASTERN CENTRE . . . *J. F. Skipsey, B.Sc.*
NORTH MIDLAND CENTRE . *J. R. Rylands, M.Sc., J.P.*
SHEFFIELD SUB-CENTRE . . . . . *F. Seddon*
NORTH-WESTERN CENTRE . *E. G. Taylor, B.Sc. (Eng.)*
NORTH LANCASHIRE SUB-CENTRE *G. K. Alston, B.Sc.(Eng.)*
NORTHERN IRELAND CENTRE . . *G. H. Moir, J.P.*
SCOTTISH CENTRE . . . . . *R. H. Dean, B.Sc.Tech.*
NORTH SCOTLAND SUB-CENTRE . . . . *P. Philip*
SOUTH MIDLAND CENTRE . *Capt. J. H. Patterson, R.A.*
RUGBY SUB-CENTRE . . . . . *P. G. Ross, B.Sc.*
SOUTHERN CENTRE . . . . . *G. D. Arden*
WESTERN CENTRE (BRISTOL) . . . *A. H. McQueen*
WESTERN CENTRE (CARDIFF) . . . *David J. Thomas*
WEST WALES (SWANSEA) SUB-CENTRE. *O. J. Mayo*
SOUTH-WESTERN SUB-CENTRE . . *W. E. Johnson*

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.